عنوان پروژه

فشرده سازی فایلهای تصویری MPEG

نام دانشجو: سارا پوری

شماره دانشجویی: ۷۹۱۱۶۲۱۰۱۸

E-mail:SP_ Sara 2000 @ Yahoo .com

چکیده: در این مقاله مروری بر استانداردهای MPEG شده و در مورد روشهای فشرده سازی فایلهای تصویری و استانداردهای کدینگ Mpeg بحث می‌شود

کلمات کلیدی: trans - image proccessing - Differential pulse code moodulation

form coding –object base coding

فهرست مطالب

عنوان	صفحه
	۱
فصل دوم: مروری بر استانداردهای MPEG	۵
۲-۱- روشهای فشرده سازی اطلاعات ویدئویی	۶
۲-۲- استاندارد MPEG –4	۶
۲-۲-۱- نمایش کد شده یك شئ چند رسانه	۹
۲-۲-۲- ترکیب اشیاء برای تولید یك صحنه	۱۰
۲-۲-۳- توصیف کردن و سنکرون کردن رشته‌های اطلاعات مربوط به یك شئ	۱۱
۲-۲-۴- انتقال رشته اطلاعات	۱۲
۲-۲-۵- تعامل با اشیاء ویدئویی	۱۳
۲-۲-۶- تعامل و مدیریت حق مولف	۱۳
۲-۲-۷- قسمت DMIF	۱۳
۲-۲-۸- قسمت سیستم	۱۴
۲-۲-۹- قسمت تصویری	۱۵
۲-۳- فایلهای MPEG-4	۱۷
۲-۴- MPEG-J	۱۸
فصل سوم: بررسی روشهای کدینگ در استاندارد MPEG-4	
۳-۱- ساختار کلی قسمت ویدئوی	۲۱
۳-۱-۱- ساختار اطلاعات	۲۳
۳-۲- بررسی کد کننده	۲۴
۳-۳- کد کردن اطلاعات مربوط به شکل یك VOP	۲۶
۳-۳-۱- انتخاب روش	۲۷
۳-۳-۲- تخمین و جبران سازی حرکت	۲۸
۳-۳-۳- کاهش دادن اندازه اطلاعات شکل	۳۰
۳-۳-۴- کد کردن بلوکهای دو دویی آلفا	۳۲

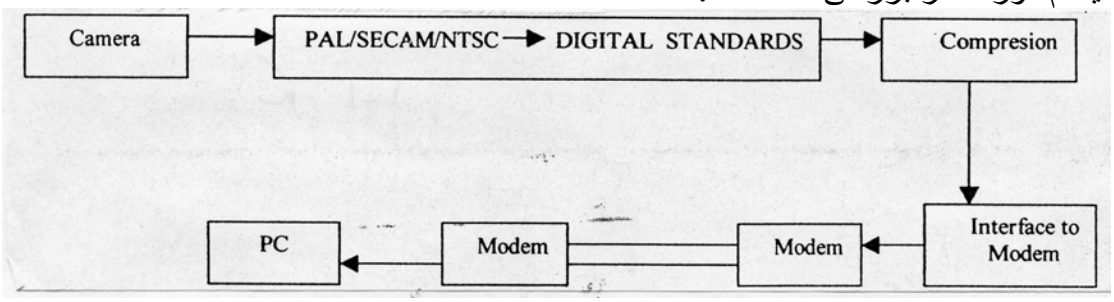
۳-۳-۵- کد کردن نقشه خاکستری آلفا ۳۳
۳-۴- روشهای تخمین حرکت ۳۴
۳-۴-۱- پر کردن ماکروبلوکهای چند ضلعی ۳۴
۳-۴-۲- انطباق ماکروبلوکهای چند ضلعی ۳۶
۳-۴-۳- روشهای جبران سازی و تخمین حرکت بصورت نا محدود ۳۹
۳-۴-۴- استفاده از روشهای پیشگویی پیشرفته ۴۰
۳-۵- واکد کننده ۴۰
۳-۵-۱- واکد کردن اطلاعات شکل ۴۱
۳-۵-۲- پیش گویی زمان اطلاعات ۴۴
۳-۵-۳- مفاهیم اصلی واکد کردن اطلاعات مقیاس پذیر ۴۵
۳-۵-۴- تولید صحنه نهایی ۴۶

مقدمه

نمایش مناسب و بهینه تصاویر ویدئویی، موضوع بسیاری از تحقیقات سالهای اخیر بوده است و با کاردهای روبه گسترش سرویسهای ویدئویی و چند رسانه‌ای مانند کنفرانس ویدئویی ، سیستمهای پزشکی از راه دور ، آموزش از راه دور و کاربردههای مشابه با آن، نیاز به گسترش فن آوری **کدینگ** سیگنالهای ویدئویی کاملاً احساس می‌شود.

در گذشته‌ای نه چندان دور رفته رفته روشهای دیجیتال به دلیل مزایایی که نسبت به روشهای آنالوگ داشت جایگزین روشهای آنالوگ شد. مهمترین ویژگیهای روشهای دیجیتال عبارتند از پردازش سریعتر و راحت‌تر نسبت به سیگنالهای آنالوگ ، حفظ کیفیت بعد از مدت طولانی ذخیره سازی و مقاومت در برابر نویز، اما مهمترین مشکل این روش پنهای باند قابل ملاحظه‌ای است که بایستی در اختیار یك سیگنال دیجیتال قرار گیرد.

با توجه نبه سیستم دریاگرام کلی آن در شکل زیر نشان داده شده است ، هر بلوك با توجه به سیستم مورد نظر بررسی شده است.



شکل ۱-۱- دیاگرام کلی سیستم

بررسی دوربینهای موجود:

دوربینهای با خروجی PAL /NTSC یا secam بدلیل استفاده از این استانداردها در ایران میتوانند مورد استفاده قرار گیرند . با توجه به استفاده از سیستم ADSL در خط انتقال و اینکه دوربین برای گرفتن تصویر از خط تولید می‌باشد، بنابر این برای داشتن کیفیت بالای تصویر از دوربینهای موجود ، دوربینهای شرکت SOVY مدل EVI-D31 OR EVI-D100 مد نظر است که توانایی فرستادن بیش از $10^{frame}/_{s}$ را دارا می‌باشند که در بخش ضمیمه detasheet آنها آورده شده است.

روشهای فشرده سازی اطلاعات

به طور کلی روشهای فشرده سازی را میتوان به دو دسته تقسیم کرد.

۱- فشرده سازی بدون خطا

۲- فشرده سازی همراه با خطا

دسته اول – فشرده سازی بودن خطا

در این روش اطلاعات بازیابی شده پس از فشرده شدن دقیقاً همان اطلاعات اولیه است. در این نوع فشرده سازی بدون توجه به خواص و نوع دیتا و نیز بدون در نظر گرفتن رابطه و وابستگی اطلاعات به یکدیگر اطلاعات را فشرده می کنیم. این روش فشرده سازی بیشتر در مورد داده‌هایی که دارای ارزش زیادی هستند استفاده می‌شود و هدف بازیابی کامل و صحیح اطلاعات اولیه است. مانند فشرده سازی Ranlength و لغت نامه‌ای ، PCM . همافمن، کدگزار حسابی، تبدیل و...

دسته دوم – فشرده سازی همراه با خطا

در این روش فشرده سازی دیتاها پس از بازیابی ممکن است دقیقاً همان مقادیر اولیه خود را بدست نیاورند. هدف کم نمودن تعداد بیت‌های اطلاعات هر چند که منجر به از دست دادن چند بیت شود.

در این نوع فشرده سازی از رابطة بین اطلاعات و خواص موجود در دیتاها و همچنین درجه اهمیت اطلاعات استفاده‌های مهم آن فشرده سازی تصاویر می‌باشد . بدلیل وابستگی اطلاعات و خطای چشم فشرده سازی همراه با خطا در تصاویر مشکلی ایجاد نمی‌کند.

انواع فشرده سازیهای همراه با خطا

این نوع فشرده سازی نیز به دو دسته تقسیم می‌گردد دسته اول کد پیش بینی کننده (predictive coding) نامیده می‌شود و به روشهایی گفته می‌شود که از خواص تصویر (روابط منطقی بین نقاط تصویر که از جمله فاکتورهای آن قابل پیش گویی بودن نقاط مجاور در تصویر و صاف بودن و یکنواختی دیتاها اشاره کرد) بهره برداری می‌کنند وتصویر را فشرده می‌کنند. تکنیکهایی نظیر

Differential pulse code Modulation در این دسته قرار می‌گیرند.

دسته دوم که کد تبدیل کننده (Transform coding) نامیده می‌شوند . عمل فشرده سازی توسط تبدیلی صورت می‌گیرد که آرایة جدیدی از تصویر می‌سازد به طوریکه تعداد زیادی از اطلاعات را در تعداد کمی نمونه نگهداری می‌کند. دو تبدیل از تبدیلهای مهم تصویر تبدیل فوریه و تبدیل کسینوس می‌باشد. علت انتخاب این دو تبدیل این است که تبدیل فومریه یکی از عمومی ترین و رایج ترین تبدیلهای موجود است و تبدیل کسینوس برای تصاویر بهترین تبدیل

به شمار می‌آید یعنی نتایج تبدیل کسینوس برای تصاویر به نتایج تبدیل اپتیمم kl بسیار نزدیك می‌باشد.

تبدیل فوریه

تبدیل فوریه برای یك رشته $\{n = 0,....,N-1\} \therefore u(n)$ به صورت زیر تعریف می‌گردد.

$$V(K) = \sum_{n=0}^{N-1} u(n) w_N^{kn} \qquad \text{k} = 0,1,....\text{N - 1} \qquad , \text{w}_\text{N} = \exp\left\{\frac{-jz\pi}{N}\right\}$$

و تبدیل عکس فوریه از رابطة زیر بدست می‌آید.

$$u(n) = \frac{1}{N} \sum_{k=0}^{N-1} V(k) w_N^{kn}$$

فرمولهای بالا خاصیت یکتایی تبدیل را بر آورده نمی کنند بنابر این در تبدیلهای تصویر از فرمولهای زیر استفاده می‌کنند.

$$U(k) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} u(k) w_N^{kn} \qquad \text{k} = 0,1,.....\text{N - 1}$$

$$U(k) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} u(k) w_N^{kn} \qquad \text{k} = 0,1,.....\text{N - 1}$$

تبدیل کسینوس

تبدیل کسینوس برای یك رشتة $\{0 \le n \le n-1\} \therefore u(n)$ به صورت زیر تعریف می‌شود

$$V(k) = a(k) \sum_{n=0}^{N-1} u(n) \cos\left[\frac{\pi(2n+1)}{2n} k\right] \qquad 0 \le \text{k} \le \text{N - 1}$$

که در آن $a(0) = \frac{1}{\sqrt{N}}$ و $0 \le \text{k} \le \text{N - 1}$ $\qquad a(n) = \sqrt{\frac{2}{N}}$

و عکس تبدیل کسینوس از رابطة زیر بدست می‌آید

$$U(k) = \sum_{n=0}^{N-1} u(k) V(k) \cos\left[\frac{\pi(2n+1)}{2n} k\right] \qquad 0 \le \text{k} \le \text{N - 1}$$

در طول کاربرد روی تبدیل کسینوس مشکلی روی تبدیل عکس کسینوسی به وجود می‌آید و آن افت سیگنال به ازای افزایش n است . برای رفع این مشکل در تبدیل عکس کسینوس از فرمول زیر استفاده می‌کنیم.

$$U(n)=\sum a'(k)V(k^{o})\cos\left[\frac{\pi(2n+0.1)}{2N}k\right] \qquad 0 \le k \le N-1$$

برای فشرده سازی با این روش یك ماتریس n* n از تصویر را جدا کرده بدین شکل که بدلیل همبستگی زیاد اطلاعات آن را به صورت رفت و برگشتی جدا می‌کنیم تا بیشترین وابستگی را در المانهای موجود در تصویر داشته باشیم . سپس در آرایه می ریزیم و تبدیل کسینوس را روی آن اعمال می‌کنیم . بسته به تعداد نقاط تبدیل، فقط چند دیتای اولیة کدرا در نظر می‌گیریم و بقیه را صفر و بدین ترتیب به جای اینکه یك ماتریس n*n از تصویر انتقال دهیم فقط K المان اول از کد را می‌توان انتقال داد یعنی اطلاعات به اندازه $\frac{n \times n}{k}$ فشرده شده و انتقال می‌یابد.

از جمله روشهای فشرده سازی همراه با خطا می‌توان کد گزار ADPCM ، کوانتیزاسیون برداری (VQ),Wavelet کد کردن به روش بلوك object-besed coding , bct ، قطعه بندی وفقی، فرکتال و .... نام برد که با ترکیب مناسبی از این روشها برای سیستم مورد نظر از استاندارد MPEG-4 استفاده شده است.

بنابراین تصویر به وسیله عناصر حساس به نور همچون CCD به سیگنال الکتریکی تبدیل می شود و سپس در مرحله نمونه برداری و کوانتیزاسیون به سیگنالهای دیجیتالی استاندارد تبدیل می‌شوند و بعد از فشرده سازی از طریق مودم ارسال می‌شود. تحقیقات لازم انجام شده است که در بخش اول (source) که شامل انتخاب دوربین و نمونه برداری از سیگنالهای خروجی می‌باشد، با توجه حساس بودن این سیستم که باید دارای وضوح و کیفیت بالا باشد نیاز به یك دوربی مناسب که دارای چرخش و کنترل از راه دور باشدانتخاب شده است(مدل انتخابی EVI-D100 شرکت SONY) در بخش دوم که انتقال اطلاعات می‌باشد با توجه به اینکه تصویر برای کنترل از راه دور می‌باشد و باید تصویری طبیعی در گیرنده داشته باشیم، پس باید با سرعت و کیفیت بالایی انتقال داده شود که با تحقیقات انجام شده در روشهای انتقال دیتا با نرخ بالا روش ADSL برای این منظور مناسب می‌باشد. (البته شرط استفاده از ADSL در خطوط تلفن قدیمی (POTS) این است که خط مورد نظر باید Leased Line باشد و switch نداشته باشد).

در بخش سوم که فشرده سازی تصویر می‌باشد با تحقیقات انجام شده روی روشهای فشرده سازی روش استاندارد MPEG-4 انتخاب شد. (البته با توجه به اینکه سرعت نرخ اطلاعات $ADSL \approx 2^{Mbit/s} \approx 1.5$ می‌باشد و با این نرخ ضریبت فشرده سازی بالایی احتیاج نداریم می‌توان از MPEG1-2 نیز استفاده نمود اما بدلیل تحقیقاتی بودن پروژه و کیفیت بالای MPEG-4 و زمینه پیاده سازی آن این روش انتخاب شد.)

با توجه به انتخاب این روشها هر کدام به صورت جزئی تر و با توجه به فهرست موجود شرح داده شده است.

فصل دوم:

مروری بر استانداردهای MPEG

استانداردهای MPEG

اولین استاندارد این گروه MPEG-1 یا ISO/IEC-11172 در سال 1992 معرفی شد. هدف عمده این استاندارد ارائه یك روش مناسب برای کد کردن سیگنالهای صوتی و تصویری به منظور ذخیره کردن این اطلاعات روی CD-ROM یا سایر درایورهای نوری می‌باشدنرخ بیت مورد استفاده کمتر از 1.5 Mb/S می باشد و برای کاربردهای صوتی از نرخهای 64Kb/S و 192Kb/s 128Kb/s را مطرح می‌کند.

روشهای مورد استفاده در این استاندارد مشابه روشهای موجود در استاندارد H. 261 می‌باشد بعد از استاندارد MPEG-1 دومین استاندارد با نام MPEG-2 یا ISO/IEC-13818 در سال 1994 معرفی شد و هدف عمد ه این استاندارد فراهم آوردن استانداردی به منظور استفاده برای کاربردهای تلویزیونی بود. استاندارد MPEG-2 کیفیتی بین تلویزیونهای NTSC/PAL و استانداردهای HDTV یا CCIR601 را ارائه می‌دهد نرخ بیت مورد استفاده در این استاندارد حدود 2-10 Mb/s می‌باشد . در واقع استانداردMPEG-2 به نیازهای روبه گسترش تلویزیونیهای دیجیتالی کابلی و سرویسهای مبتنی بر شبکه از طریق مد انتقال غیر همزمان (ATM) کاربردهای ضبط دیجیتال ویدئویی (VTR) و گسترش شبکه های پخش تلویزیونی دیجیتال از طریق ماهواره و شبکه های زمینی پاسخ می‌دهد. در این روش طیف وسیعی از روشهای مختلف و پارامترهای متفاوت در نظر گرفته شده اند.

مثلاً قاب تصویر و نرخ بیت آن توسط کاربر قابل تنظیم است و این استاندارد تنها ساختار رشته داده و ساختار واکد کننده را معین می کند به همین دلیل این امکان را برای صنایع مختلف ایجاد کرده است که بتوانند کیفیت سیگنال تلویزیونی را منطبق با نرخ بیت مورد استفاده تنظیم کنند.[2]

۲-۱- روشهای فشرده سازی اطلاعات ویدوئویی.

به طور کلی اصل فشرده سازی تصویر مبتنی بر کاهش افزونگی (Redundancuy) موجود در تصاویر می‌باشد . به طور کلی دو نوع افزونگی در اطلاعات ویدئویی وجود دارد

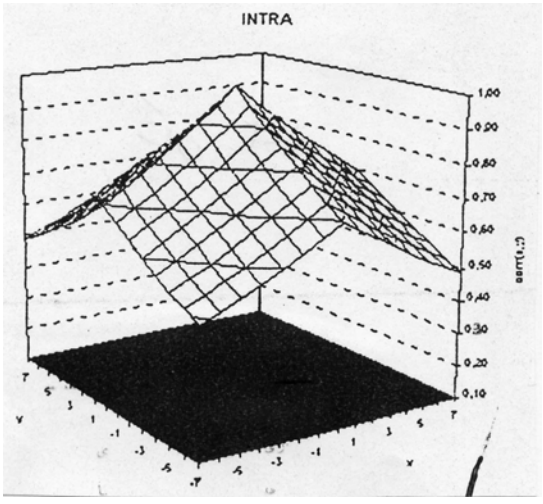
۱-افزونگی آماری (Statistical redandancy)

۲- افزونگی مشاهده گر (Subjective redundancy)

و هدف نهایی کاهش دادن هر دو نوع افزونگی به منظور رسیدن به کمترین اطلاعات برای کد کردن می‌باشد که منتهی به رسیدن به پایین ترین نرخ بیت مورد نیاز می‌شود.

علاوه بر کاهش دادن افزونگی که به معنی کاهش میزان اطلاعات قابل ارسال میباشد بایستی با استفاده از روشهای کدینگ آنتروپی (Entropy ciding Technique) به میزان کمترین نرخ بیت نیز دست یافت. به طور کلی اساس روشهای کدینگ در استانداردهای MPEG بر کاهش افزونگی آماری تصویر مبتنی است که شامل افزونگی در حوزة مکان یا حوزة زمان می‌شود.

افزونگی در حوزة زمان یعنی اینکه مقدار روشنایی یا رنگ در یك نقطه مشخص را می‌توان با استفاده از مقدارهای متناظر در قابهای بعدی یا قبلی تصویر بدست آورد و افزونگی در واحد مکان به این معنی است که اطلاعات روشنایی یا رنگ در یك نقطه تصویر بسیار شبیه اطلاعات متناظر در همسایگی آن نقطه در همان قاب تصویر می‌باشد، به طور مثال در شکل زیر که تابع همبستگی Cor(x,y) محاسبه شده است مشاهده می شود که مقدار سیگنال در یك نقطه شباهت زیادی به همسایگی آن دارد و به همین دلیل می‌توان مقدار سیگنال در یك نقطه را از روی اطلاعات همسایگی آن بدست آورد به منظور کاهش میزان افرونگی مکانی در یك قاب تصویر از روشهایی استفاده می‌شود که به آن کدینگ درون قابی (Interframe) گفته می‌شود که در واقع همبستگی (Correlation) مکانی در یك قاب تصویر را کاهش می‌دهد از جمله روشهایی که میزان همبستگی مکانی را در یك قاب تصویر کاهش می‌دهد روشهای تبدیل کدینگ مثل تبدیل گسسته سینوسی یا DCT (Discrete Cosine Transform) می‌باشد.



شکل ۲-۱ : تابع همبستگی یك نقطه در قاب تصویر

به علاوه به منظور کاهش دادن افزونگی زمانی در بین چندین قاب متوالی نیز از روشهایی مانند DPCM یا مدولاسیون کدپاس تفاضلی (Differntial Pluse Code Modulation) استفاده

می‌شود که در واقع نوعی پیش گویی اطلاعات در یك نقطه از روی اطلاعات قاب قبلی و با استفاده از مقدارهای جبران سازی خطا می‌باشد.

در استانداردهای سری MPEG از یك ترکیب بهینه (Hybrid DPCM/DCT) که ترکیب هر دو این روشها است استفاده می‌شود یعنی ابتدا از DPCM به منظور کاهش دادن افزونگی زمانی و سپس از روش DCT به منظور کاهش افزونگی مکانی استفاده می‌شود.[1]

استاندارد MPEG-4

در ابتدای امر این استاندارد به منظور فراهم آوردن روشی برای کد کردن سیگنالهای تلویزیونی روی نرخ بیتهای خیلی پایین، در حد 64kb/s که توسط شبکه های اینترنتی و خطوط (Local Area Network) فراهم می‌شود مطرح شد اما بعدها هدف این استاندارد تغییر کرده تبدیل به روشی برای کد کرد سیگنالهای تلویزیونی بر اساس محتوی این سیگنال ها شد.

یکی از مزایای MPEG-4 این است که بر خلاف استانداردهایی که قبلاً مطرح شده‌اند و هیچ کدام امکان دستکاری اطلاعات در مرحله‌ای که هنوز این اطلاعات بازسازی نشده‌اند را فراهم نمی‌کنند. در این استاندارد با یك روش کدینگ متناسب با محتوای (Content-based coding) این امکان را فراهم میکند که قبل از باز سازی تصویر امکان دستکاری آن را داشته باشیم. به علاوه سیستم بینایی انسان(Human visual system) یا HVS اطلاعات تصویر را با توجه شکل و حرکت و سطح آنها دریافت می کند و روشهای مقیاس پذیری قبلی که عبارت بودند از مکان و زمان هیچ کدام منطبق با سیستم بینایی انسان نیست در حالیکه با فراهم آوردن امکاناتی مانند مقیاس پذیری تصویر بر اساس اشیاء موجود در صحنه (VOP Scalcbitity) این امر محقق می‌شود.

این استاندارد در سال ۱۹۹۵ آغاز شد و در سال ۱۹۹۸ پیش نویس آن نهایی شد . برای نرخ بیتهای در حد 64 kb/s مطرح شد اما در حال حاضر تا 4Mb/s را نیز شامل می‌شود که برای کاربردهای موبایل و شبکه PSTN (Public Switch Telephone Network) نرخ بیتی معادل 5-64lbps را استفاده می کند و برای کاربردهای تلویزیونی نرخ بیتی معادل 46lb/s-2Mb/s را استفاده میکند. کیفیتی بین QCIF(Quartert Common Intermediate Format) و HDTV را دارد و عمدتاً برای کاربردهای مخابرات چند رسانه‌ای از طریق شبکه اینترنت، بازیهای کامپیوتری ، شبکه های چند رسانه ای متحرك ، کاربردهای مانند تلفن تصویری و کنفرانس ویدئویی اشاره کرد.

به طور کلی استاندارد MPEG-4 برای به تحقق رساندن اهداف زیر مطرح شده است:

۱- معرفی مفاهیم جدیدی به نام شی رسانه ای (Media object) : این اشیاء رسانه ای می‌توانند واحدهای صوتی و یا تصویری و یا صوتی و تصویری باشند که تشکیل یك واحد مجزا به نام شی تصویری را می‌توان هم توسط یك دوربین ضبط کرد و هم اینکه مانند گرافیك کامپیوتری توسط کامپیوتر تولید کرد.

۲- امکان ترکیب اشیاء مختلف چند رسانه‌ای برای تولید یك صحنه صوتی تصویری.

۳- امکان مالتی پلکس کردن و سنکرون کردن دیتاهای مربوط به اشیاء چند رسانه‌ای، تا اینکه امکان ارسال این اطلاعات روی کانالهای شبکه و تخصیص کیفیت متناسب با سرویس مورد نظر فراهم شود.

٤- امکان دستکاری و تغییر صحنه صوتی و تصویری تولید شده در سمت گیرنده

این استاندارد با عنوان ISO/IEC –14496 مطرح شده است و دارای شش قسمت می‌باشد، این قسمتها عبارتند از قسمت سیستم ، قسمت ویدئویی، قسمت صوتی، قسمت عملکرد، قسمت نرم افزار و قسمت ساختار لایة انتقال که در ادامه به توضیح این قسمتها خواهیم پرداخت.[4]

۲-۲-۱- نمایش کد شده یك شئ چند رسانه ای

(Cdede representation of a media object) همانطوریکه در شکل مشاهده می‌شود یك صحنة ویدئویی که توسط MPEG-4 کد میشود می تواند شامل چندین نوع مختلف از اشیاء صوتی و تصویری باشد که به صورت ارتباط یك درخت با اجزایش در نظر بگیریم برگهای این درخت را می‌توان اشیاء صوتی و تصویری (Audio video object) یا AVO بنامیم که این اشیاء می‌توانند یکی از انواع زیر باشند:

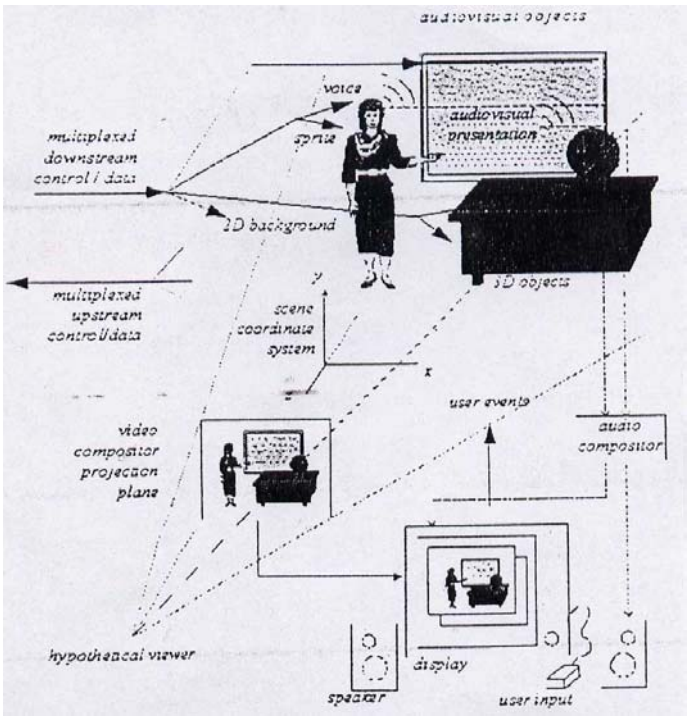
-تصاویر ثابت مثل صحنه یك تصویر

- اشیاء ویدئویی مثل یك گوینده بدون پشت صحنة آن.
- اشیاء صوتی مثل صوت مربوط به این گوینده

البته این امکان در استاندارد MPEG-4 وجود دارد که بتوان اشیاء طبیعی و مصنوعی (گرافیك ساخته شده توسط کامپیوتر) را باهم ترکیب کردو در یك صحنه نشان داد که حتی این اشیاء میتوانند دو یا سه بعدی نیز باشند به همین دلیل به علاوه اشیاء بالا می‌توانیم اشیاء زیر را داشته باشیم:

- متن و گرافیك
- صوت و سر مصنوعی گرافیکی و متن مربوط به آن که برای ساختن و تولید گفتار و متحرك سازی سر استفاده می‌شود.
- صدای مصنوعی

به طور کلی هر شئ چه صوتی و چه تصویری ، دارای یکسری اطلاعات است که به ما اجازه می‌دهد این شئ را در یك صحنه مورد استفاده قرار می‌دهیم که این اطلاعات یکسری اطلاعات توصیفی می‌باشد. در حقیقت یك شئ در یك صحنه مستقل از اطرافش و پشت صحنه کد می‌شود . این روش کد کردن صحنه‌های صوتی و تصویری بهترین روش برای فراهم آوردن امکاناتی است که MPEG-4 در نظر دارد. از جمله این امکانات می‌توان مقاومت در برابر خطا، بیرون کشیدن و یا نوشتن (Edit) یك شئ و یا فراهم آوردن یك شئ با چندین لایة مختلف مقیاس پذیری را ذکر کرد.



شکل ۲-۲- ساختار درختی یك صحنه از تصویر

۲-۲-۲- ترکیب اشیاء برای تولید یك صحنه (Composition of media object)

در شکل قبل نحوة ترکیب شدن اشیاء مختلف صوتی و تصویری به منظور تولید یك صحنة کامل نشان داده میشود به صور کلی دو نوع شئ در یك صحنه وجود دارند.

۱- اشیاء اولیه (primitive media object)

۲- اشیاء ترکیبی(compound media object)

در حقیقت هر شئ اولیه را می‌توان یکی از برگهای درخت سلسله مراتبی تصویر در نظر گرفت و از ترکیب چندین شئ اولیه یك شئی ترکیبی تولید می‌شود که می‌توان آن را به صورت یك شاخه از این درخت در نظر گرفت به عنوان مثال ترکیب صوت گوینده و تصویر آن که هر دو یك شئ اولیه هستند تولید شئ ترکیبی گوینده را می کند.

چنین دسته بندی هایی این امکان را فراهم می‌سازد که بتوانیم یك صحنه پیچیده را طراحی کنیم و بعلاوه بتوانیم یك مفهوم معنی دار به یك شئ اختصاص دهیم که امکان ایجاد تغییر در صحنه را برای کار بر فراهم سازد.

امکاناتی که با توجه به این نوع طراحی صحنه فراهم می‌شود عبارتند از :

- قرار دادن یك شئ در هر محلی در مختصات داده شده
- اعمال تبدیلاتی به منظور تغییر دادن ظاهر مکانی یك شئ
- دسته بندی یك سری اشیاء اولیه برای تشکیل دادن یك شئ ترکیبی
- تغییر دادن مسیر دید یا جهت پخش صدا از دید کار بر
- اختصاص دادن یك رشتهٔ دیتا به اشیاء صوتی یا تصویری به منظور توصیف کردن ویژگیهای آن مانند صوت ، بافت جابه جا شده یك شئ پارامترهای متحرك سازی یك صوت گرافیکی و مانند آن.

البته باید توجه داشت که ساختار اولیه روشهای مورد استفاده از VRML (Vitrual Reality Modeling Language) گرفته شده است.

۲-۲-۳- توصیف کردن و سنکرون کردن رشته های اطلاعات مربوط به یك شئ.

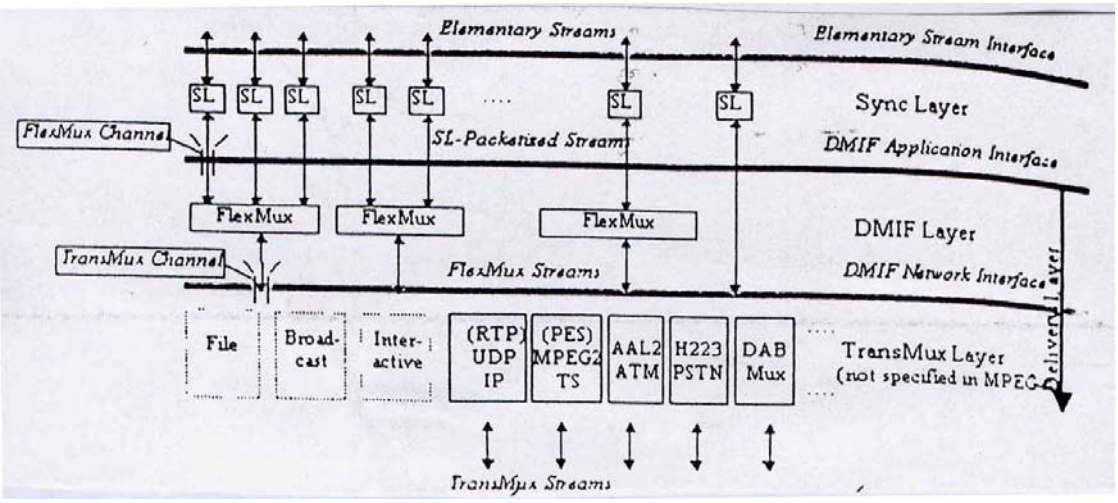
(Description and synchronization of streaming data for a media)

یك شئ رسانه‌ای ممکن است به یك رشته اطلاعات نیاز داشته باشد که خود این رشته اطلاعات در یك یا چند رشته اطلاعات دیگر حمل می‌شود و در واقع یك توصیف کننده (Descriptor) همه رشته‌هایی که مربوط به یك شئ رسانه‌ای می‌باشد را معین می‌سازد.

علاوه بر این، توصیف کننده می‌توانند فوق داده (Meta data) را نیز در خود داشته باشد که این فوق داده در حقیق همان اطلاعات مربوط به محتوی فایل از جمله اندازهٔ فایل و سازندهٔ فایل و نوع اطلاعات موجود در فایل و ... باشد. بعلاوه حق انحصاری استفاده از این اطلاعات را نیز میتواند شامل شود (کپی رایت). در عین حال هر کدام از این رشته های اولیه (Elementry stream) نیز داری یك سری توصیف کننده می‌باشند که در آنها اطلاعاتی مثل اینکه نیاز به چه واکد کننده‌ای در مقصد دارد و یا اینکه برای انتقال نیاز به چه کیفیت انتقال یا

(Quality of Service) دارد (شامل نرخ بیت ماکزیمم، نرخ خطای حداکثر و ... می باشد) را منتقل می‌کند.

عملیات سنکرون سازی با عمل بر چسب زمانی زدن روی واخدهای دسترسی (Access unit) داخل این رشته های اولیه صورت می‌گیرد و در حقیقت لایه سنکرون سازی وظیفه یکسان سازی این واحدها ی دسترسی را بر عهده دارد. يك واحد دسترسی می‌تواند يك قاب صوتی یا تصویری باشد.



شکل ۲-۳- لایة انتقال و لایة سنکرون سازی

۲-۲-۴- انتقال رشتة اطلاعات (Delivery of streaming data)

انتقال اطلاعات از منبع به مقصد به صورت سنکرون شده توسط در ولایة سنکرون سازی و لایة انتقال انجام می‌پذیرد خود این لایة انتقال دارای دو لایه می‌باشد.

در لایه اول انتقال ،ES هایی که دارای QOS مشابه هستند با یکدیگر مالتی پلکس می‌شوند که این عمل در حقیقت توسط Flex Mux (Flexible multipiexer) انجام می‌شود و به منظور استفاده بهینه از امکانات شبکه صورت می‌پذیرد.

لایة بعدی یا Trans Mux (Transport Multiplexer) در واقع لایه‌ای است که متناسب QOS مورد در خواست خدمات انتقال را ارائه میدهد.

البته در استاندارد MPEG-4 فقط مدار واسط با این لایه معرفی شده است در حالیکه پیاده سازی حقیقی اطلاعات و سیگنالهای کنترلی بایستی متناسب با یك پروتکل انتقال صورت بگیرد.

هر نوع پروتکل انتقال مثل RTP/UDP/IP یا AAL5/ATM و یا MPEG-2 می‌توانند به عنوان یك پروتکل انتقال در نظر گرفته شوند و انتخاب نوع این پروتکل بستگی به کار بر و فراهم آورندة سرویس (Service provider) دارد این قابلیت این اجازه را به ما می‌دهد که بتوانیم در رنج وسیعی از کاربردها از MPEG-4 استفاده کنیم.

استفاده از لایة Flex MUX اختیاری است و می‌تواند حذف شود در این صورت لایة Trans MUX وظیفه دارد که همه امکانات مورد نظر را تأمین کند و وجود آن اجباری است لایه سنکرون سازی نیز حتماً بایستی موجود باشد با توجه به این طرح امکانات زیر برای ما فراهم شده است.

- می‌توانیم واحدهای دسترسی را مشخص کنیم، اطلاعات مربوط به زمان و اطلاعات مربوط به کلاك مرجع را ارسال کنیم و میزان تلفات اطلاعات (Data loss) را نیز معین کنیم.
- می‌توانیم به صورت دلخواه اطلاعات را از ES های مختلف گرفته و داخل رشته های Flex MUX بگذاریم.
- QOS مورد نیاز برای یك ES را در هر Flex MUX تعیین کنیم.
- نیاز به چنین QOS را به منابع حقیقی شبکه تبدیل کند.
- ES را به اشیاء رسانه‌ای مربوط ربط دهد.
- اطلاعات ES را که در Flex MUX و Trans MUX قراردارند حمل کند.

۲-۲-۵- تعامل با اشیاء ویدئویی (Interaction with media objects)

در حقیقت یك کاربر یك صحنة متناسب با آنچه طراح (Author) برایش در نظر گرفته مشاهده می‌کند و بسته به اینکه چقدر این طراح برای کاربر آزادی عمل در نظر گرفته باشد کار بر میتواند با صحنة نمایش ارتباط داشته باشد. از جمله:

- نقطه دیدن و یا شنیدن را با حرکت در صفحه تغییر دهد.
- یك شئ را از یك نقطه معین به جای دیگری منتقل کند.
- ترتیب و توالی اتفاقات را با کلیك کردن روی یك شئ خاص تغییر دهد.

- در صورتیکه امکانات صوتی چیدن زبانه فراهم شده باشد یك زبان خاص را انتخاب کند.

البته این امکان وجود دارد که طراحی های پیچیده تری انجام شود مثلاً یك تلفن مجازی زنگ بزند و با کاربر صحبت کند.

۲-۲-۶- تعامل و مدیریت حق مولف

(Management and Interaction of Intellectual property) این امکان که حق استفاده از یك محصول خاص تنها در اختیار خالق قرار داشته باشد و یا به اصطلاح حق کپی رایت محفوظ باقی بماند بسیار مورد توجه قرار دارد به همین دلیل در استاندارد MPEG-4 سعی شده است که ابزارها و دستورات لازم برای این منظور فراهم شود. پیچیدگی و جزئیات این امکان تحت عنوان بالا (قبل) معرفی شده است. در واقع MPEG-4 حق کپی رایت را با ذخیره یك سری مشخص کننده های منحصر به فرد (Unigue Identifier) که توسط یك سیستم بین‌المللی شماره گذاری ، مشخص می‌شود و حق استفاده از یك محصول خاص را معین می‌کند حفظ می‌کند البته به دلیل اینکه همة محصولات ممکن است دارای این عدد نباشند این استاندارد این امکان را فراهم می‌کند که هر محصول را بتوانیم با یك کلید نیز مشخص کنیم.

۲-۲-۷- قسمت DMIF (Delivery Multimedia Interaction Framework) لایه DMTF وظیفه آوردن امکانات زیر را بر عهده دارد.

- فراهم آوردن یك مدار واسط (DMIF) یکسان برای هر کار بر صرفه نظر از اینکه این ارتباط با یك طوج از راه دور و یا از طریق انتشار و یا توسط یك وسیله ضبط اطلاعات انجام می‌شود
- کنترل پیاده سازی کانالهای Flex Mux
- فراهم آوردن امکان استفاده از شبکه های یکسان مثل (Internet protocol) IP , ATM ، موبایل ، PSTN , ISDN باند باریك بین دو زوج کاربر.

۲-۲-۸- قسمت سیستم

در واقع MPEG-4 یك روش پیشرفته برای فشرده سازی اطلاعات صوت و تصویر می‌باشد همانطوریکه میدانیم اطلاعات کد شده که به صورت رشته های دیتا یا ES ها در آمده اند میتوانند به صورت جداگانه ذخیره و یا ارسال شوند به همین دلیل برای اینکه صحنة نهایی را در سمت دیگر تولید کنند بایستی با یکدیگر ترکیب شوند وظیفة ترکیب این اطلاعات صوتی و تصویری توسط قسمت سیستم صورت می‌پذیرد.

به عبارت دیگر قسمت سیستم در این استاندارد ارتباط بین مولفه‌های صوتی و تصویری که یك صحنه را تشکیل می‌دهند مشخص می‌کند و این ارتباط را در دو سطح توصیف می‌کند.

۱- BIFS (Binary Format For Scenes) یا حالت با نیری صحنه:

که در واقع ارتباط بین اشیاء در صحنه از نظر زمانی و مکانی را معین می‌کند. ممکن است این امکان در اختیار کار بر قرار داده شده باشد که بتوانند ترتیب قرار گیری اشیاء مختلف را در یك صحنه دوباره مرتب کنند و جهت دیدشان را در یك محیط مجازی سه بعدی تغییر دهند که این عملیات از طریق دستورات معرفی شده انجام می‌پذیرد.

۲-ODS (Object Descriptors) یا توصیف گرهای اشیاء:

در یك مرحله پایین تر از BIFS ، این OD ها ارتباط بین هر شئ و ES های مربوط به آن را نشان میدهد به عنوان ثال صوت و تصویر مربوط به یك شرکت کننده در یك سیستم کنفرانس ویدئویی توسط توصیف گرهای مربوط به آن به یکدیگر مربوط می‌شوند.علاوه بر این OD ها اطلاعات اضافی دیگری مانند ویژگیهای یك واکد کننده برای اینکه این اطلاعات را جداسازی (Parse) کند، آدرس URL که برای دسترسی به ES ها مورد نیاز است، حق انحصاری استفاده از این اطلاعات و سایر موارد را با خود حمل میکند.

علاوه بر دو قابلیت مهمی که در بالا توسط BIFS و OD فراهم شده است امکانات دیگری نیز در این سیستم MPEG-4 قرار داده شده است از آن جمله:

- **تعامل (Interactivity) :**تعامل بین سرویس دهنده (Server) و سرویس گیرنده (Client) تنظیم کردن و مسیر دادن به ارتباط بین اشائ در یك صحنه متناسب با وقابع و یا عکس العملهای کاربر ، تنظیم کردن و مسیر دادن به عکس العملهای کار بر از طریق یك مدل عمومی
- **Flexmux :**این قسمت این امکان را فراهم می‌سازد که رشته های اطلاعات از منابع مختلف را در یك رشته نهایی جا دهیم و اطلاعات زمانی مربوط به آنها را نیز منتقل کنیم.
- فراهم آوردن ابزاری برای ذخیره کردن اطلاعات MPEG-4 در فایلهایی از نوع Mp4 (MPEG-4 File format)

ارتباط بر قرار کردن با انواع متفاوتی از پایایه‌ها و شبکه ها با استفاده از دستورات Java API که تحت عنوان MPEG-J مطرح می‌شود.

استقلال لایة انتقال از پشته‌های (stack) مربوط به پروتکل انتقال مورد نظر ، مثل RTP/VDP/IP و یا MPEG-2

- نمایش متن با حمایت از زبانهای بین‌المللی، امکان انتخاب نوع فونت، زمان بندی و سنکرون سازی آن
- مقدار دهی اولیه و مدیریت مداوم روی بافرهای ترمینال گیرنده
- مکانیزم های بازیابی اطلاعات ، سنکرون ساختن آنها و تعیین زمان بندی بین اطلاعات.
- فراهم آوردن امکان شناسایی حق انحصاری استفاده از این اطلاعات صوتی و تصویری با استفاده از منابع داده (Data sets)

۲-۲-۹- قسمت تصویری (visual)

همانطوریکه می دانیم استاندارد MPEG-4 این امکان را فراهم می‌کند که تصاویر واقعی یا مصنوعی را در کنار همدیگر فشرده سازی کنیم به طور مثال این امکان را به ما می‌دهد که در یك کنفرانس ویدئویی یك شرکت کنندة مجازی داشته باشیم. که به صورت اجمالی امکاناتی که توسط قسمت تصویری این استاندارد برای ما فراهم می‌کند گفته شده است:

-فرصت سیگنالهای ویدئویی.

۱-نرخ بیتی در حد 5kb/s-10Mb/s

۲-سیگنالهای ویدئویی چه از نوع ترتیبی (Progressive) و چه از نوع هم بافته (Interlaced) را می‌تواند استفاده کند تا امکان کمتری تولید کند به طور مثال برای تصاویر ثابت می‌توان تا ۱۱ سطح با تفکیك پذیری مختلف داشت و در مورد سیگنالهای ویدئویی می‌توان تا ۳ سطح با تفکیك پذیری مختلف داشت.

۲-مقیاس پذیری زمانی : (Temporal scalability) :

این امکان را به واکد کننده می‌دهد که با باز سازی بخشی از رشته داده میزان تفکیك پذییری زمانی تصویر را کاهش دهد.

۳- مقیاس پذیری کیفی (Quality scalability)

این امکان را فراهم می‌کند که بتوانیم رشتة داده را به تعدادی از رشته‌های داده با نرخ بیتهای مختلف تقسیم کنیم به گونه ای که هنوز ترکیب آنها با یکدیگر تولید یك رشته ها معنی دار کند. عملیات جداسازی این رشته‌های داده می تواند در زمان انتقال در واکد کننده صورت پذیرد و کیفیت نهایی وابسته به این است که چه تعداد از این لایه‌ها برای باز سازی تصاویر استفاده شده‌اند.

- **کد کردن اطلاعات مربوط به شکل و کانال انتقال (shape and channel coding) .**

به طور کلی کد کردن اطلاعات شکل به منظور فراهم آوردن تسهیلاتی در توصیف اشیاء ویدئویی با ابعاد و شکل دلخواه فراهم شده است. این روش با استفاده از نقشة دو دوئی (Binary alpha map) و یا نقشة خاکستری (Gray scalr alpha map) انجام می‌پذیرد. در حالت دودوئی ، این نقشه تنها معین می‌کند که کدامیك از این نقاط تصویر متعلق به یك شئ هست و می‌تواند دو حالت صفر و یك داشته باشند. در حالیکه در روش سطوح خاکستری این امکان فراهم شده است که رکوشنایی هر نقطه از شکل نیز مشخص شود.

-مقاومت در برابر خطا در محیطهای مستعد خطا

(Roboustness in error prone enuironments)

امکاناتی در این استاندارد گذاشته شده است که در محیطهایی مانند شبکه های بی سیم که درای پهنای باند کم کو احتمال خطای بالا می‌باشد اطلاعات ویدئویی را منتقل کند.

-متحرك سازی صورت (Face Animation)

در قسمتی از این استاندارد امکان ارسال پارامترهایی که به منظور تنظیم و متحرك سازی یك صورت مصنوعی به کار می‌رود قرار داده شده است. البته خود این مدلها توسط MPEG-4 استاندارد نشده است تنها پارامترهای آنها توسط MPEG-4 معرفی شده است که این پارامترها عبارتند از FDP و FAP

FDP: پارامترهای توصیفی صورت (Facial Definition Parameter) نامیده می‌شوند.

FAP : پارامترهای متحرك سازی صورت (Facial Animation Parameter) نامیده می‌شوند.

در گیرنده متناسب با اطلاعات FAP که دریافت می‌شود حرکاتی از صورت مانند حرکات لب و دیگر اعضای صورت به وجود می‌آید و در همین زمان اطلاعات FDP نیز دریافت شده که ظاهر این صورت را از یك حالت کلی و عمومی تبدیل به یك وضعیت ویژه با شکل مخصوص به خودش و بافت مخصوص خودش می کند.

- **کد کردن مشهای دو بعدی:**

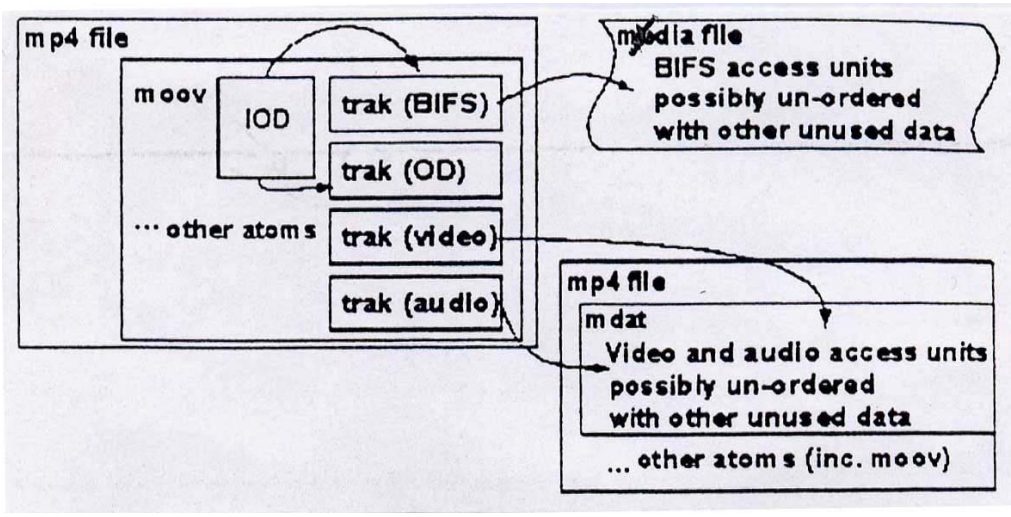
به علاوة روشهای قبلی که برای کد کردن یك شکل با ابعاد دلخواه در نظر گرفته می شود که مستلزم استفاده از نقشة دودوئی و یا نقشة خاکستری شکل بود روش مش نیز وجود دارد که به منظور کد کردن یك شکل با ابعاد دلخواه از طریق تقسیم کردن این شئ به مجموعه ای از مثلث و مربع صورت می‌پذییرد البته در روش MPEG-4 از زیر مجموعه های مثلثی استفاده می‌شود.[3,4]

۲-۳-فایلهای MPEG-4

فایلهای MP4 به منظور نگهداری اطلاعات مربوط به یك رشته دیتا از نوع MPEG-4 به گونه ای که امکان مدیریت و ویرایش این اطلاعات هنوز فراهم باشد و به علاوه بتوان این اطلاعات را روی سیستم که محتوی این اطلاعات می‌باشد و یا سیستم دیگری که از طریق شبکه به آن متصل می‌باشد نمایش طراحی شده است.

شکل و ساختار این فایلها بایستی مستقل از هر نوع پروتکل انتقال باشد و در عین حال امکان انتقال بهینه را به صورت عمومی نیز فراهم سازد.

در شکل نموداری از یك مثال ساده اراده شده است که حاوی سه رشته اطلاعات می‌باشد یك فایل MP4 از ساختارهای مبتنی بر شئ که اتم (Atom) نامیده می‌شود تشکیل شده است و هر اتم را توسط یك بر چسب و یك طول مشخص میکنیم . اکثر این اتمها در واقع حاوی اطلاعاتی از نوع فوق داده (Meta data) می‌باشند که طول و اشاره گر وسایر اطلاعات مربوط به دیتاهای اصلی را با یك ساختار سلسله مراتبی مشخص میکند.



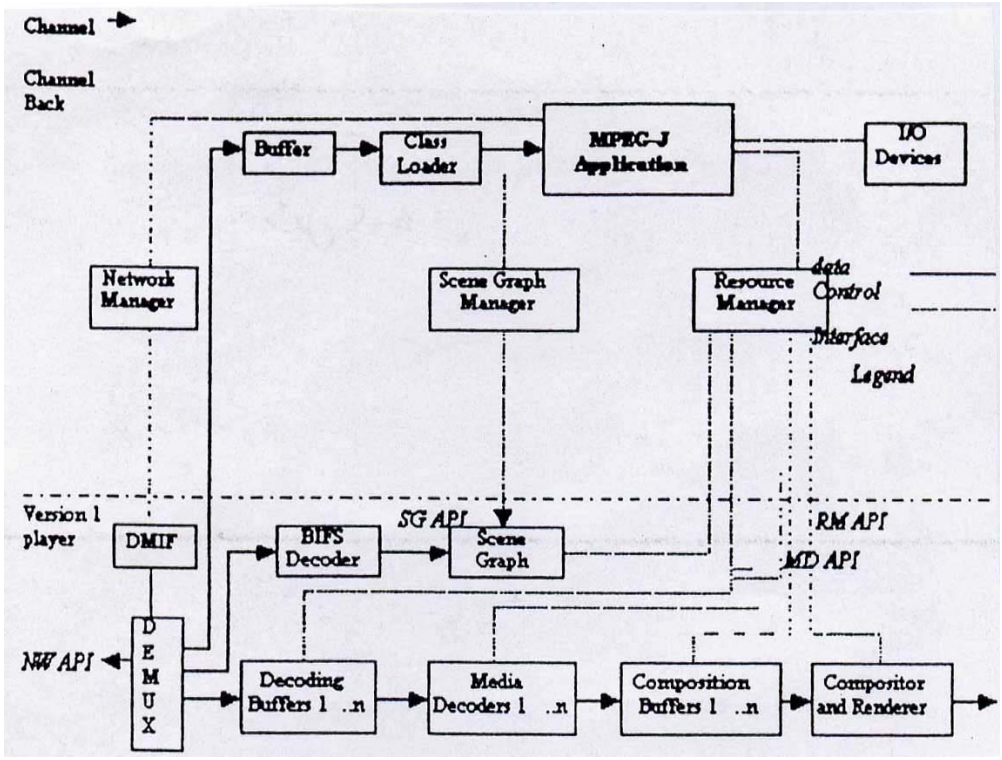
شکل ۲-۴- ساختار ذخیره یك رشته MPEG-4 روی فایل

این مجموعه در یك اتم کلی قرار گرفته است که movie atom یا moov نامیده می‌شود اطلاعات اصلی در محل دیگری قرار دارد که می‌تواند داخل همین فایل باشد و یا اینکه کلاً خارج از این فایل و در یك فایل دیگر ذخیره شده باشد. محل ذخیره دیتاهای اصلی را mdat (media data) یا اتم mdat می‌نامند.

۲-۴- MPEG-J

MPEG-J یك ساختار برنامه نویسی می‌باشد که یك سری دستورات API که به منظور کار کردن نمایش دهنده های MPEG-4 با کدهای Java مشخص می کند. با ترکیب کردن اطلاعات MPEG-4 با یکسری کد قابل اجرا اجرا این امکان برای این اطلاعات فراهم شده است که از

طریق این کدهای قابل اجرا یکسری از عملیات و پردازشهای پیچیده را کنترل کنند. شکل یك سیستم نمایش دهنده MPEG-4 را نشان می‌دهد که به موتور نمایش دهنده هم معروف است.
زیر سیستم MPEG-J که مسئول کنترل کردن موتور نمایش می‌باشد که در قسمت بالا نشان داده شده است . دستورات Java در یك رشته دادهٔ اولیه مجزا به پایانه‌های MPEG-4 ارسال می‌شود به همین دلیل می‌توان آن را مستقیماً به محیط اجرایی MPEG-J فرستاد که در این محل در واقع این برنامه دسترسی به مولفه های مختلف و اطلاعات مربوط به نمایش دهنده های MPEG-4 را دارد. دستورات مربوط به مدیریت گراف صحنه (Scene graph API) امکان دسترسی به گراف صحنه و تغییر آن را دارند مثلاً امکان بازرسی گراف، امکان اصلاح کردن گره‌ها و اطلاعات مربوط به گره و یا اضافه و کم کردن یك گره از یان گراف فراهم است.
توسط دستورات مدیریت منابع (resource manage API) قابلیت تنظیم کردن نحوه اجرا و نشان دادن این اطلاعات فراهم شده است.
دستورات کنترل واکد کننده ها (Media decoder API) امکان کنترل واکد کننده های موجود در پایانه های MPEG-4 را فراهم می‌کند.
دستورات شبکه (Network API) امکان تأثیر گذاری و ارتباط با شبکه را فراهم ساخته است
دستورات پایانه های (Terminal cability API) زمانی استفاده می‌شوند که اجرا برنامه روی ترمینالهای MPEG-4 متحمل و ممکن باشد.[2,3,5]



شکل ۲-۵- اجرای دستورات API

فصل سوم:

بررسی روشهای کدینگ در استاندارد MPEG-4

بررسی روشهای کدینگ در استاندارد MPEG-4

۳-۱- ساختار کلی قسمت ویدئویی:

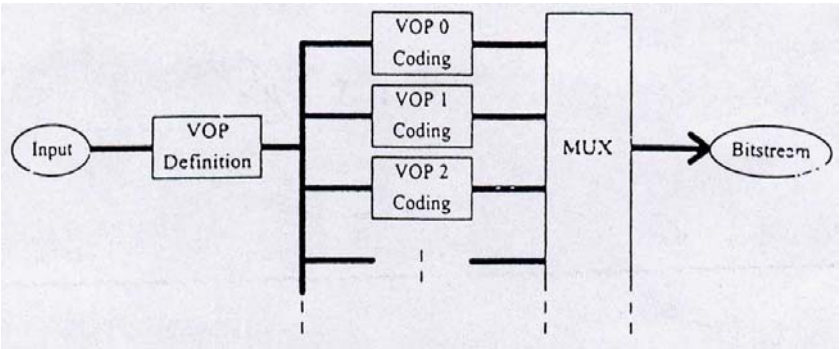
همانطور که قبلاً اشاره شد اساس استاندارد MPEG-4 مبتنی بر تبدیل یك تصویر به زیر مجموعه‌ای از اشیاء موجود در شکل و کد کردن این اشیاء به صورت جداگانه می‌باشد.

به طور کلی هر شئ که در این تصویر مشخص می‌شود به صورت یك VOP (Video object plane) معرفی می شود و اطلاعات مربوط به این شئ که به صورت متوالی کد می شود تولید یك رشته داده می‌کند و به صورت (Video Object) vo معرفی می‌شود که در واقع یك سری VOP پشت سر هم با شکل و موقعیت زمانی متفاوت در قابهای مختلف تصویر هستند.

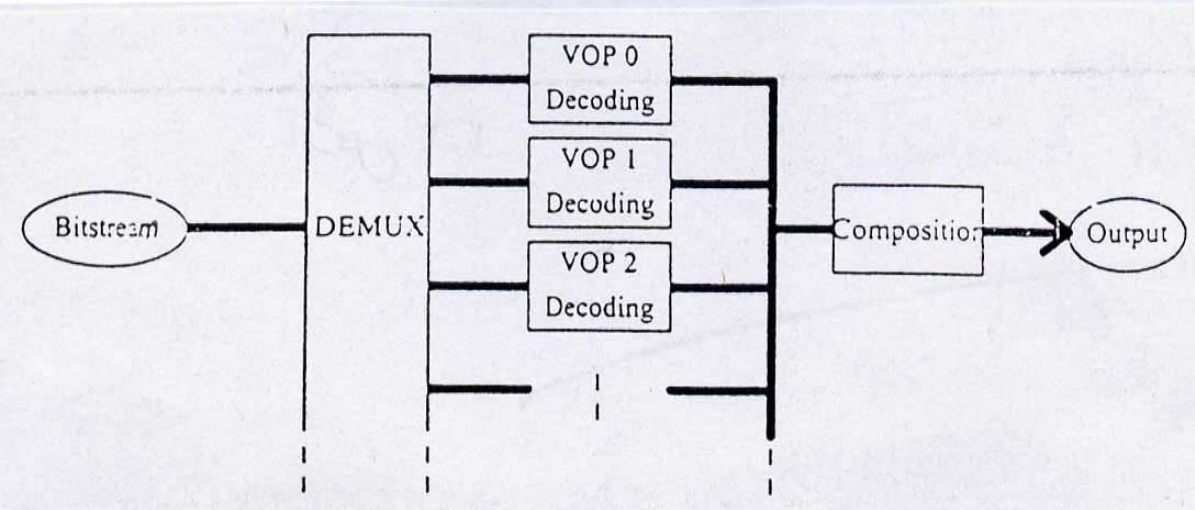
حرکت و شکل و بافت مربوط به یك VOP که تشکیل یك VO را میدهد در لایه‌های جداگانه به نام VOL (Video Object Layer) کد می‌شوند.

مهمترین مزایای استفاده از این روش فراهم آوردن روشی برای فشرده سازی بهینه تصویر می‌باشد زیرا این امکان را فراهم میکند که در صورتیکه قسمتی از تصویر اهمیت چندانی ندارد نرخ بیت کمتری به آن اختصاص دهیم و قسمتهای از تصویر که دارای اهمیت بالاتری است با نرخ بیت بالاتری کد می‌شود . به علاوه امکاناتی مانند مقیاس پذیری زمانی و مکانی تصویر رانیز فراهم می‌سازد.

همانطوریکه میدانیم در این استاندارد هر صیسحنه توسط اشیاء موجود در صحنه معرفی میشود. البته به منظور هماهنگ کردن این استاندارد با روشهای قبلی این امکان وجود دارد که بتوانیم یك قاب را به صورت یك شئ مستطیلی شکل در نظر بگیریم شکل زیر (در صفحة بعد) یك کد کننده ساده را نشان می‌دهد.



شکل ۳-۱ نمودار بلوکی کد کننده



شکل ۳-۲- نمودار بلوکی واکد کننده

مهمترین قسمت این کدر (Encoder) بخشی است که یك صحنه را تبدیل به شئ ویدئویی می‌کند. معمولاً هر شئ بایستی یك مفهوم معنی دار باشد و معمولاً یك کار بر و یا یك الگوریتم هوشمند می‌توانند شئ های مختلفی را با پارامترهای مختلف انتخاب کند . البته اگر شئ ویدئویی مصنوعی توسط کامپیوتر خلق شده باشد در این صورت مرز این شئ و کل تصویر بسیار واضح می‌باشد و یا می‌توان برخی از اشیاء درون یك تصویر را با استفاده از روشهای کلید های رنگی (Chroma Keying) از بقیه تصویر مجزا کرد، البته این مختص اشیایی است که به روش فیلمبرداری جلوی صفحهٔ آبی به کل صحنه اضافه شده اند و با روشهای تجزیه و ترکیب رنگ می‌توان مرز این اشیاء را کاملاً مشخص کرد و یا اینکه با استفاده از برخی روشهای نمیه اتوماتیك (Semi –Outomatic) این اشیاء را مرز بندی کرد.

بخث تشکیل یك شء ویدئویی در استاندارد MPEG-4 طراحی نشده است و یك کد کننده MPEG-4 فرض می کند که یك سری VOP با مرزهای مشخص ، اطلاعاتی را برای ارسال دارند.

به منظور انطباق روشهای فشرده سازی مورد استفاده در این استاندارد با آن چه در قبل مورد استفاده بودذه است، هر شئ ویدئویی فشرده شده دارای دو دسته اطلاعات می‌باشد اول اطلاعات شکل ای شئ و دوم اطلاعات مربوط به رنگ و روشنایی نقاط داخل این شکل، که این اطلاعات به صورت مولفه‌های YUV تعریف می‌شود و در قابهای متوالی با فواصل زمانی که از قبل تعریف شده است ارسال می‌شوند.

در زمان کد کردن اطلاعات کار بر می‌تواند روی نحوهٔ کد کردن اطلاعات تأثیر بگذارد به عنوان مصال در مرحله کنترل کد کردن (Coding Control) برخی از اشیاء راکد نکند و یا

نرخ بیت معینی را به برخی از این اشیاء اختصاص دهد به علاوه در مرحله مالتی پلکس کردن می‌تواند مراحل و نحوه ترکیب شدن VOP ها را تغییر دهد.

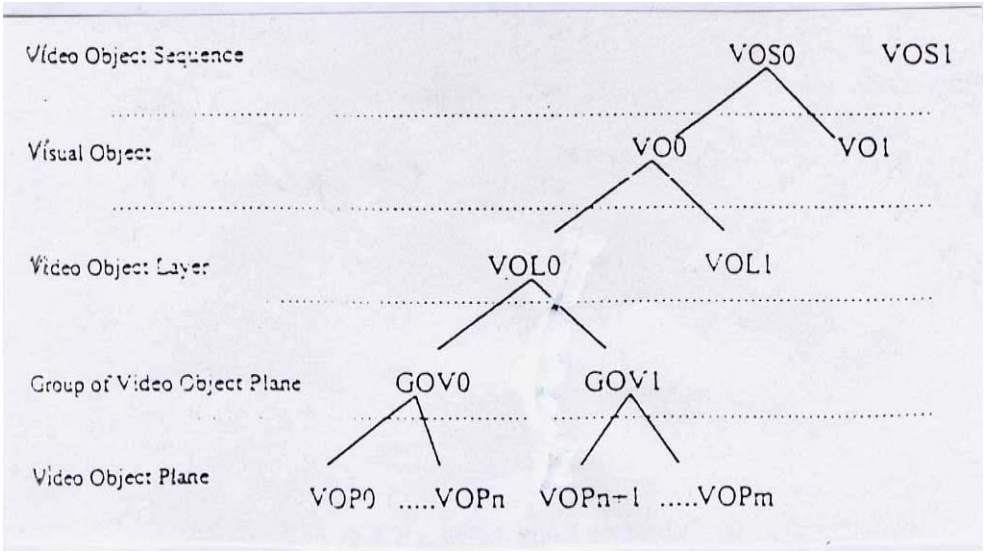
در شکل ۳-۲ نمایش بلوکی یك واکد کننده نشان داده شده است که دقیقاً عملیات معکوس را انجام می‌دهد و از ترکیب این اشیاء با یکدیگر صحنهٔ نهایی را تولید می‌کند. در این مرحله نیز کاربر می‌تواند روی تصویر نهایی تأثیر بگذارد مثلاً در مرحله دی مالتی پلکس کردن با در خواست پردازش قسمتی از یك رشتهٔ داده همهٔ اطلاعات را دریافت نکند و یا در مرحلهٔ ترکیب این VO ها با یکدیگر روی موقعیت و یا نحوه ترکیب شدن آنها در صحنه تأثیر بگذارد.[5]

۳-۱-۱- ساختار اطلاعات

این نمودار شکل سلسله مراتبی اطلاعات را در این روش مبتنی بر شئ (Object-baded) نشان می‌دهد . به طور خلاصه هر تکه ویدئویی (video session) یا VS تشکیل شده است.

تعدادی شئ که هر کدام از آنها در یك یا چند لایه که VOL (Video Object Layer) نامیده می‌شوند منتقل شده اند و هر VOL در واقع رشتهٔ زمانی از VOP ها می‌باشد.

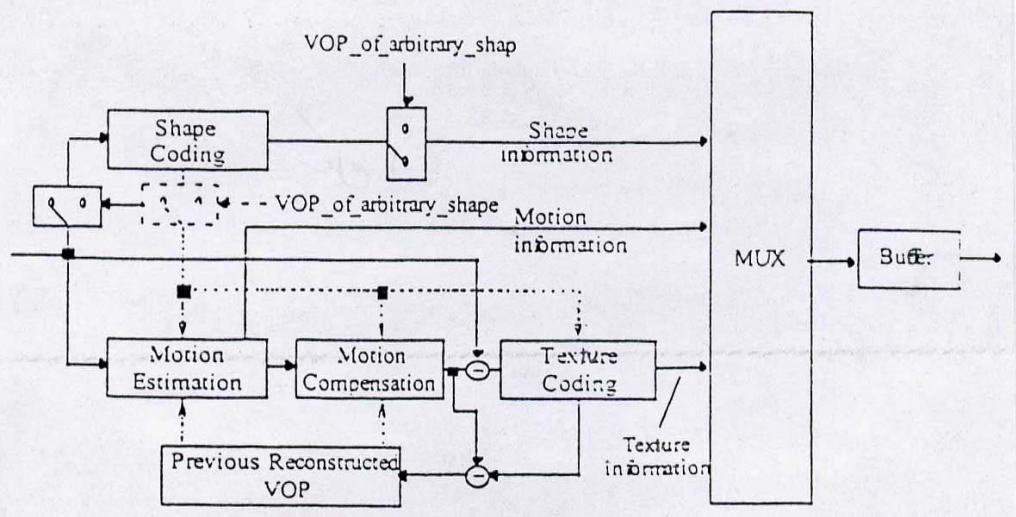
یك شئ ویدئویی میتوانند دارای یك شکل دلخواه باشد. می‌تواند یك شئ دو بعدی یا سه بعدی طبیعی یا مصنوعی باشد. در حالت تك لایه‌ای در واقع تنها یك لایه برای هر شئ در نظر گرفته شده است.



شکل ۳-۳- ساختار رشته اطلاعات

۳-۲- بررسی کد کننده

مطابق شکل زیر یك کد کننده از دو قسمت تشکیل شده است. قسمت اول مربوط به کد کردن اطلاعات شکل یك VOP می‌باشد . در حالیکه قسمت دوم ، اطلاعات مربوط به رنگ و روشنایی را در این VOP کد می‌کند، هر VOP توسط یك چارچوب مرزی (Bounding Rectangle) مشخص می شود.



شکل ۳-۴- ساختار کد کننده VOP

الگوریتم هایی که در کد کردن بافت شکل (Texture) استفاده می‌شود همان الگوریتمهای مورد استفاده در استانداردهای MPEG1/2 می‌باشند ، یعنی تبدیلات DCT .DPCM و عملیات جبران سازی حرکت و بردارهای حرکت با توجه به اینکه اساس این روشها مبتنی بر ساختار ماکرو بلوکی می‌باشد به همین دلیل در کد کردن هر VOP نیز ابتدا بایستی VOP اطلاعات تبدیل به بلوکهای مجزای ۱۶×۱۶ شود و سپس پردازش روی این ماکروبلوکها انجام شود با توجه به شکل دلخواه یك VOP ممکن است در تبدیل این VOP به واحدهای ماکرو بلوکی همه قسمتهای این ماکرو بلوك حاوی اطلاعات نباشد یعنی تولید ماکرو بلوك چند ضلعی (Modified یا Polygon) می‌کند به منظور تبدیل این چند ضلعی ها به ماکروبلوکهای عادی از برخی روشهای پرکردن (Padding) استفاده می‌شود. به طور کلی اطلاعات شکل یك VOP می‌تواند با دو روش مختلف منتقل شود.

۱- نقشه دودوئی آلفا (Binary alpha plane)

۲- نقشهٔ خاکستری آلفا (Gray scale alpha plane)

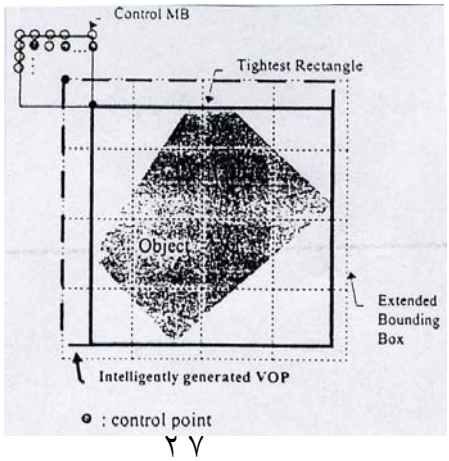
در نقشه دودئی ، اطلاعات مربوط به شکل با صفر و یك ردن نقطهٔ متناظر در این چارچوب مرزی تولید می‌شود که صفر بودن این نقطه نشان می‌دهد که این نقطه متعلق به VOP جاری

می‌باشد در نقشه خاکستری به جای اختصاص دادن صفر و یك ، عددی در فاصله صفر تا ۲۵۵ منتقل می شود که این عدد علاوه بر حضور یا عدم حضور VOP در آن نقطه، قسمتی از اطلاعات روشنایی نیز منتقل می‌کند. به این اطلاعات نقشه بیتی (Bitmap) نیز گفته می‌شود.

برای هر VOP که در یك قاب از تصویر تعریف می‌شود چارچوب مرزی کوچکترین قاب مستطیلی در نظر گرفته می‌شود که اضلاع آن مضارب صحیح از ۱۶ باشند برای تعیین کوچکترین چارچوب مرزی از یك روش هوشمند مطابق مراحل زیر استفاده می‌شود . با این روش حداقل تعداد ماکرو بلوك در چار چوب مرزی قرار می‌گیرد.

ابتدا برای هر VOP کوچکترین مستطیلی که همهٔ VOP را بپوشناند به گونه ای در نظر گرفته می‌شود که مختصات نقطه بالا و سمت چپ آن زوج باشد اگر مختصات شروع قاب تصویر باشد به این معنی است که یك VOP مسیطیلی در نظر گرفته شده و هر قاب تصویر یك VOP در نظر گرفته شده است و نیازی به ارسال اطالاعات مربوط به شکل نمی‌باشد و به همین دلیل از مرحلهٔ کد کردن اطلاعات شکل صرفه نظر می‌کنیم.

در غیر اینصورت یك ماکرو بلوك کنترل در گوشهٔ بالا و سمت چپ این مستطیل مطابق شکل زیر تشکیل می‌دهیم . هر نقطه با مختصات زوج روی این ماکروبلوك را میتوان به عنوان نقطه کنترلی در نظر بگیریم سپس با در نظر گرفتن این نقطه به عنوان مرجع یك چارچوب مرزی را به سمت راست و پایین تولدی کنیم به گونه‌ای که طول و عرض این چارچوب مضربی از ۱۶ باشد و همه VOP را نیز بپوشاند. بعد از تولید هر چارچوب مرزی تعداد ماکروبلوکهایی را که حداقل روی یك پیکسل (PIXEL) اطلاعات داشته باشند را می‌شماریم. این عملیات را برای همه نقاط کنترلی داخل این ماکروبلوك انجام می‌شود. از بین همه این نقاط ، نقطه ای که کمترین تعداد ماکرو بلوك را به ازاء چارچوب مرزی رسم شده از این نقطه تولید می‌کند را به عنوان مرجع نهایی انتخاب می‌کنیم. مطابق با این روش چارچوب مرزیی را انتخاب می‌کنیم که دارای حداقل تعداد ماکرو بلوك می‌باشد و در عین حال کل این VOP را نیز می‌پوشاند.



شکل ۳-۵- تشکیل هوشمند VOP

از روی نقشهٔ بیتی مربوط به اطلاعات روشنائی می‌توان نقشه بیتی اطلاعات رنگ را نیز تولید کرد. این عملیات توسط یك روش زیر نمونه برداری (Sub sampling) از نقشه بیتی اطلاعات روشنایی انجام می‌پذیرد.

اگر از نقشه دودوئی آلفا استفاده شده باشد برای یك همسایگی ۲×۲ در صورتیکه حداقل یك نقطه در این همسایگی مقدار ۲۵۵ را داشته باشد مقدار نهایی ۲۵۵ خواهد بود و اگر از نقشه خاکستری آلفا استفاده شده باشد مقدار آلفا روی نقشه اطلاعات رنگ، میانگین رنگ چهار نقطه مجاور می‌باشد.

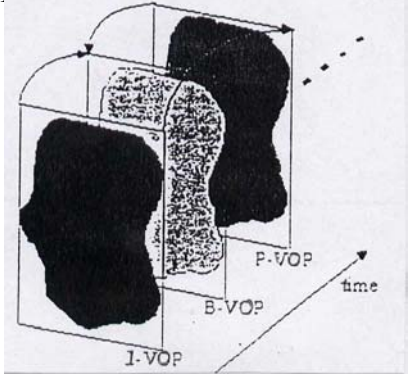
۳-۳- کد کردن اطلاعات مربوط به شکل یك VOP

در این قسمت روشهای تعریف شده برای کد کردن اطلاعات دودوئی و یا سطوح خاکستر که برای انتقال اطلاعات مربوط به شکل استفاده می‌شود را بررسی می‌کنیم.

هر نوع اطلاعات شکل از این پس به نام نقشه های آلفا (Alpha plane) معرفی می‌شود . برای کد کردن نقشه های دودوئی آلفا از روش CAE (control- based-Arithmetic Encoding) استفاده می‌شود در حالیکه برای کد کردن نقشه خاکستری آلفا از روشهای DCT همراه با روشهای جبران سازی خطا استفاده می‌شود.

می‌دانیم هر چارچوب مرزی مضرب صحیحی از بلوکهای ۱۶×۱۶ می‌باشد و عملیات کد کردن و واکد کردن روی این بلوکها که بلوکهای آلفا نامیده می‌شود انجام می‌شود.

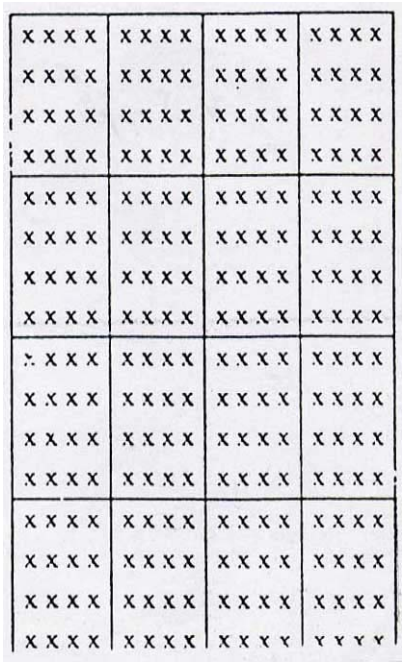
در صورتیکه همه نقاط داخل یك بلوك آلفا مقدار صفر داشته باشند قبل از انجام هر گونه پردازشی از آن صرف نظر می‌شود. مطابق آنچه در استانداردهای $MPEG\,1/2$ بیان شد هر قاب تصویر را می‌توانیم به سه صورت I یا P یا B کد کرد. در این استاندارد نیز از همه این روشها استفاده می‌شود با این تفاوت که به جای در نظر گرفتن قاب مستطیلی می‌توان یك شکل دلخواه که به صورت VOP تعرییف شده است را در نظر گرفت و اطلاعات یك VOP را با استفاده از کد کردن درون قابی (I-VOP) و یا بین قابی (P,B-VOP) منتقل کرد.



شکل ۳-۶- انواع مختلف VOP

در ابتدا روشهایی که برای کد کردن نقشه دودوئی آلفا استفاده می‌شود بررسی خواهیم کرد و سپس در ادامه روشهای مورد استفاده در نقشه خاکستری آلفا بررسی خواهد شد.

اطلاعات نقشه دودوئی آلفا را میتوانیم با استفاده از روشهای درون قابی (INTRA) برای حالت I-VOP و یا روش بین قابی (INTER) در P-VOP و B-VOP کد کردن و هر بلوك اطلاعات آلفا را بلوکهای دودوئی آلفا یا (Binary Alpha Block) می‌نامیم، این BAB از بلوکهای ۴×۴ مطابق شکل تشکیل شده است. به طور کلی هر BAB شامل ۱۶ بلوك پیکسل یا (Pixel Block) PB می‌باشد.



شکل ۳-۷- یك BAB متشکل از ۱۶ عدد PB می‌باشد

۳-۳-۱- انتخاب روش

در برخی کاربردها نیازی نیست که دقیقاً BAB را از روشهای بدون تلفات کد کنیم بلکه تا یك محدوده معینی تلفات قابل قبول می‌باشد که تحت عنوان کیفیت مطلوب یا ACQ (Accepted Quality) از آن یاد می‌شود متناسب با این کیفیت مطلوب نوع روش انتخابی برای کد کردن BAB متفاوت می‌باشد . بنابراین برای پیاده سازی این مفهوم پارامتری تحت عنوان آستانه آلفا (Alpha –th) معرفی می‌شود که این مقدار می‌تواند یك از مقدارهای {256,….64,32,16,0} باشند . تابع SAD-PBI (BAB,BA B' (عبارت از مجموعه قدر مطلق

تفاوت بین BA B' BAB, روی I این PB ، با توجه به تعاریف فوق تابع، ACQ (BA B') به صورت زیر تعریف میشود.

ACQ (BA B')=MIN (acq1,acq2,…,acq16)

acqi=0 if SAD-pbi>16*alpha-th

acqi=1 other wise

با انتخاب alpha-th=0 در واقع یك روش بدون تلفات را در نظر گرفته ایم . در صورتیکه با در نظر گرفتن alpha-th=256 یك روش کاملاً تلفاتی را در نظر می‌گیریم یعنی اگر همه نقاط نیز اطلاعات غلط منتقل کنند باز هم قابل قبول می‌باشد.

با تعریف تابع فوق بهترین روش برای کد کردن یك BAB از بین هفت روش موجود انتخاب می‌شود. این هفت روش به صورت خلاصخ به صورت زیر تعریف می‌شوند:

۱-"all-0" : همه مقدارهی نقاط صفر در نظر گرفته می‌شود و هنگامی اتفاق می‌افتد که ا ختلاف آن با یك BAB0 از حد آستانه کمتر باشد.

۲-"all-255" : همه مقدارهای نقاط ۲۵۵ در نظر گرفته می شود و هنگامیکه اتفاق می‌افتد که اختلاف آن با یك BAB255 از حد آستانه کمتر باشد.

۳-"Intra CAE" : با استفاده از روشهای کدینگ درون قابی و با استفاده از کد CAE بدون استفاده از تخمین حرکت کد می‌شود.

۴-"MVDs==088,NO update" : از روشهای تخمین حرکت استفاده می شود ولی مقدار خطا کد نمی‌شود به علاوه مقدار اختلاف بردار حرکت با آنچه بردار حرکت پیش گوئی شده است صفر می‌باشد.

۵-"MVDs! =0& NO Update" : از روشهای تخمین حرکت استفاده می‌شود ولی مقدار خطا کد نمی شود به علاوه مقدار اختلاف بردار حرکت با آنچه پیش گوئی شده نیز صفر نمی‌باشد و بایستی ارسال شود.

۶-"MVDs==0& inter CAE" : از روشهای تخمین حرکت استفاده می‌شود ولی مقدار خطا نیز با استفاده از روشهای CAEکد می شود به علاوه مقدار اختلاف بردار حرکت با آنچه پیش گوئی شده نیز صفر می‌باشد .

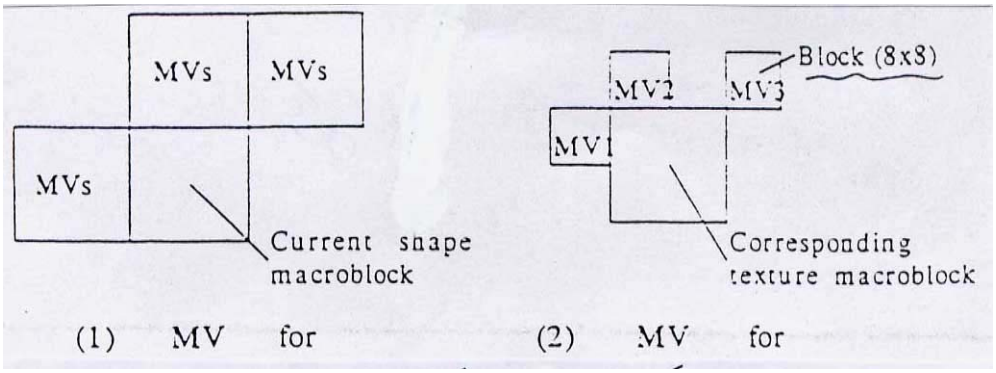
۷-"MVDs==0& inter CAE" : از روشهای تخمین حرکت استفاده می‌شود ولی مقدار خطا نیز با استفاده از روشهای CAEکد می شود به علاوه مقدار اختلاف بردار حرکت با آنچه پیش گوئی شده نیز صفر نمی‌باشد و بایستی ارسال شود.

البته واضح است که اگر در مد I-VOP باشیم تنها به روش اول ممکن است استفاده شوند.

۳-۳-۲- تخمین و جبرانسازی حرکت

برای پیاده سازی روش جبران سازی حرکت در کد کردن اطلاعات شکل دو مرحله اصلی بایستی انجام شود، ابتدا تعدادی بردار برای پیش گویی حرکت که تحت عنوان MVPS (Motion rector Preditionfor shape) معرفی می‌شوند حدس می‌زنیم.در قدم بعدی بردار حرکت واقعی را محاسبه می‌کنیم.

بردارهای پیش گویی حرکت را از دو طریق می‌توان حدس زد. اول با استفاده از بردارهای حرکت روی BAB که اخیراً کد شده اند. دوم با در نظر گرفتن بردارهای حرکت روی بلوکهای مربوط به اطلاعات روشنایی متناظر که اخیراً کد شده‌اند مطابق شکل زیر برای هر BAB می‌توان شش بردار حرکت در نظر گرفت.



شکل ۳-۸- بردارهای کاندید برای MVP_s

با درنظر گرفتن ترتیب MVs3, MVS2 , MVS1, MV3, MV2, MV1 اولین بردار حرکت معتبر از میان این بردارها به عنوان MVPs انتخاب می‌کنیم.

بعد از انتخاب MVPs برای یافتن بردار حرکت شکل یا MVs (MV for shape) مراحل زیر را طی می‌کنیم.

به ازاء هر MVPs اختلاف BAB جاری با BAB که MVPs به آن اشاره میکند را محاسبه می‌کنیم. اگر این اختلاف کمتر از حد آستانه مشخص شده باشد خود MVPs به عنوان MVs پذیرفته می‌شود. در صورتیکه شرط بالا ارضا نشود MV را در حوالی MVPs جستجو می‌کنیم ورودی یك ناحیه ۱۶-/+ در هر دو جهت افقی و عمودی عملیات جستجوی انجام می‌پذیرد ونقطه‌ای که در آن کمترین SAD تولید شده است را به عنوان MVs اختلاف آن را با MVPs ارسال خواهیم کرد که به صورت تفاضل بردار حرکت برای شکل یا MVDs تعریف می‌شود.

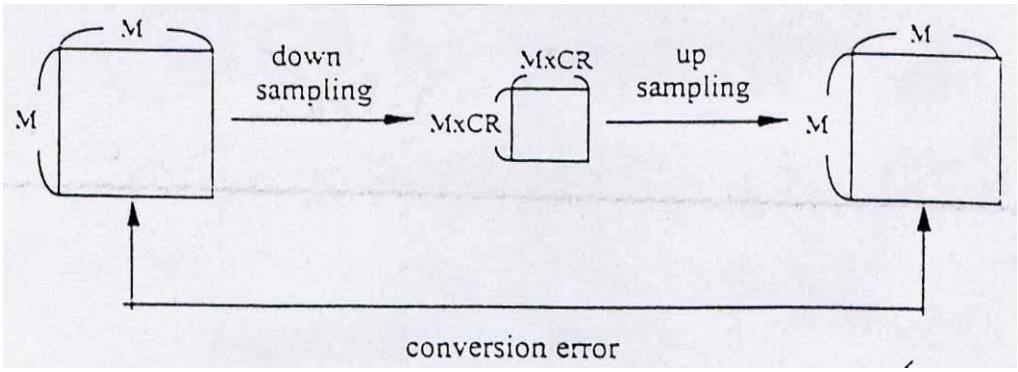
MVDs=MVs-MVPs

بعد از اینکه MVS پیدا شد عملیات جبران سازی حرکت متناسب با MVs روی هر بلوك ۱۶× ۱۶ صورت می‌گیرد و نوع جدیدی از بلوك ایجاد می‌شود که تشکیل شده است از یك بلوك BAB با ابعاد ۱۶×۱۶ که حاشیه‌ای با ضخامت یك پیکسل اطراف آن را پوشانیده است به این بلوك M C BAB گفته می‌شود. مقدار پیکسلهای این بلوك یا جابجایی ساده در راستای MV بدست می‌آید. اگر نقطه‌ای که به آن اشاره می‌شود خارج از نقشه دودوئی آلفا باشد مقدار متناظر این نقاط صفر در نظر گرفته می‌شود.

۳-۳-۳- کاهش دادن اندازه اطلاعات شکل

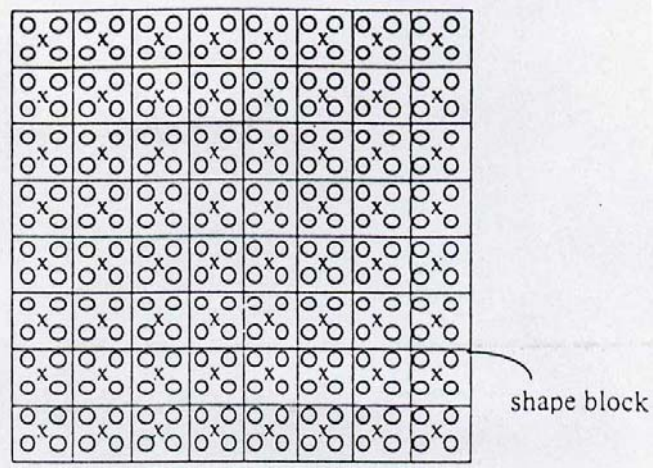
یکی از روشهای مورد استفاده برای کاهش دادن حجم اطلاعات و کنترل نرخ بیت عبارتست از گوچك کردن سایز بلوکهای آلف حین پردازش ، البته این روش برای همه نوع BAB به جز نوع "Noupdate", "all-255","all-0" قابل پیاده سازی می‌باشد.

مطابق شکل زیر برای پیدا کردن ضریب تبدیل که تحت عنوان CR (Conversion Ratio) معرفی می‌شود و می‌تواند یکی از اعداد ۱و۱/۲و۱/۴ باشد ابتدا تبدیل کاهش نمونه انجام می‌شود . سپس روی همین اطلاعات تبدیل افزایش نمونه صورت می‌گیرد و نتیجه نهایی با BAB اولیه مقایسه می‌شود در صورتیکه اختلاف کمتر از آستانه



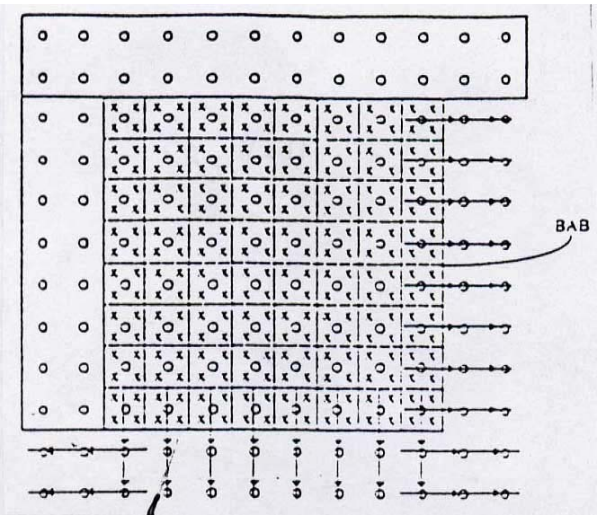
شکل ۳-۹- تعیین ضریب تبدیل در یك حلقه بسته

تعریف شده باشد همین ضریب CR مورد قبول می‌باشد در غیر این صورت این ضریب را بزرگتر می‌کنیم. اگر CR=0.5 باشد روی بلوك ۲×۲ میانگین مقدارهای روی نقاط را حساب می‌کنیم. اگر این میانگین بزرگتر از ۱۲۸ باشد مقدار نقطه نمونه برداری شده عدد ۲۵۵ و در غیر اینصورت صفر در نظر گرفته می‌شود. در صورتیکه CR=0.25 باشد عملیات تعریف شده برای 0.5 دو بار متوالی انجام می‌شود.



شکل ۳-۱۰-کاهش نمونه

در شکل بالا "O" معرف نقاط اولیه و "X" معرف نقاط نمونه برداری شده می‌باشد. برای افزایش نمونه و افزایش داده سایز اطلاعات مطابق شکل زیر بایستی از اطلاعات BAB های مجاور نیز استفاده کرد. با توجه به اینکه BAB بالا و سمت چپ BAB جاری ، از قبل موجود می‌باشد از بین مقادیر برای انجام عملیات درون یابی و یافتن مقدارهای "x" استفاده می‌شود و برای بلوکهای سمت راست و پایین آخرین مقدار روی مرز این بلوکها مطابق شکل زیر به سمت بیرون بسط داده می‌شوند.



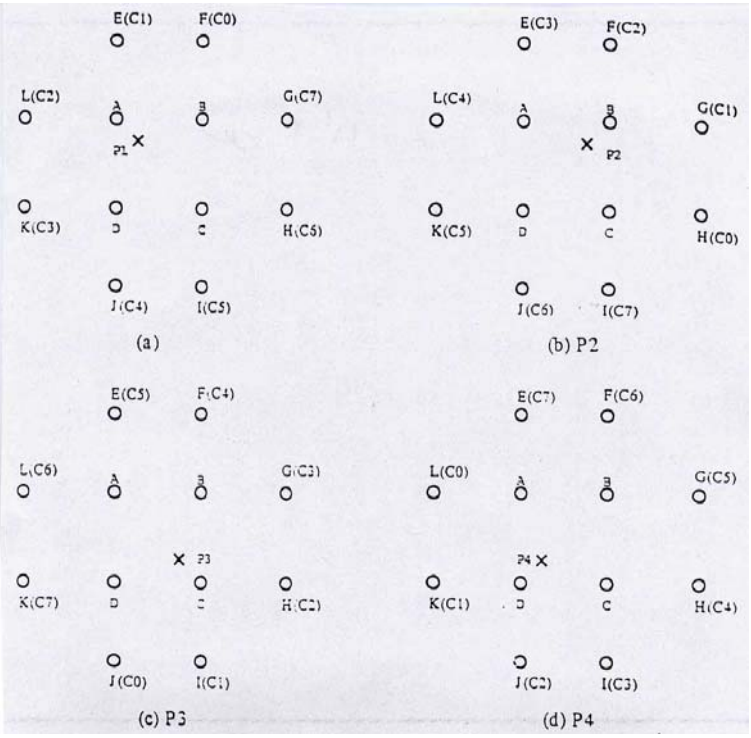
شکل ۳-۱۱- افزایش نمونه

برای انجام عملیات افزایش نمونه ابتدا عدد Cf که یك ضریب فیلتر ۸ بیتی می‌باشد تعیین می‌کنیم(C1 تا C8) . مطابق شکل مقدارهای "O" در محاورت مقدار "X" است که بایستی درون یابی شود. با داشتن $Cf = \sum_k C_k Z^k$ و با استفاده از جدول TH[Cf] که آستانه مورد نظر را به ازاء یك Cf مشخص بدست می‌دهد از روابط زیر استفاده کرده و مقدارهای P1 تا P4 را محاسبه می‌کنیم.

$$P1: if\left(4\times A+2\times\left(B+C+D\right)+\left(E+F+H+I+J+K+L\right)>Th[cf]\right)then'1'else'0'$$

$$P2: if\left(4\times B+2\times\left(A+C+D\right)+\left(E+F+H+I+J+K+L\right)>Th[Cf]\right)then'1'else'0'$$
$$P3: if\left(4\times C+2\times\left(B+A+D\right)+\left(E+F+H+I+J+K+L\right)>Th[Cf]\right)then'1'else'0'$$
$$P4: if\left(4\times D+2\times\left(B+C+A\right)+\left(E+F+H+I+J+K+L\right)>Th[Cf]\right)then'1'else'0'$$

در واقع این روش دقیقاً عکس عمل کاهش نمونه می‌باشد که در حالتهای مختلف حل شده است و به صورت چهار معادلهٔ بالا و یك جدول Th[cf] پیاده سازی شده است.



شکل ۳-۱۲- درون یابی و شروط درون یابی

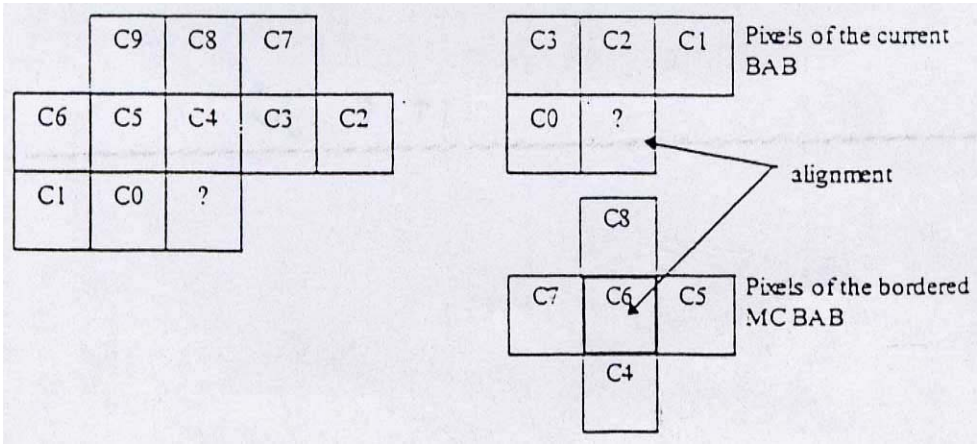
در صور تیکه یك BAB بعد از انجام عملیات نمونه برداری تبدیل به بلوك تمام صفر و تمام یك شد ودر تمام زیر بلوکها خطای محاسبه شده کمتر از آستانه تعریف شده بود در این صورت این بلوك به صورت "all0-0", "all0-255" کد می‌شود . بعد از انجام عملیات تبدیل سایز یك BAB می تواند به ازای CR=1 اندازه ۱۶×۱۶ و به ازای CR=0.5 اندازه ۸×۸ و به ازای CR=0.25 اندازه ۴×۴ داشته باشد.

۳-۳-۴- کد کردن بلوکهای دودوئی آنها

اطلاعات در BAB با استفاده از روشهای CAE کد می‌شوند. روش کد گزاری CAE دو روش عمده دارد.

Intra CAE و intraer CAE روش intra برای I-VOP و P-VOP استفاده می‌شود و از روش inter CAE در P-VOP, B-VOP استفاده کرد.

روش CAE برای کد کردنه هر نقطه داخل یك BAB استفاده می‌شود و این نقاط به صورت پویش افقی (Raster) کد می‌شوند مراحل پردازشی که برای کد کردن یك نقطه بایستی استفاده شود به شرح زیر می‌باشد ، ابتدا مقدار عدد زمینه (context) را متناسب با نوع کدینگ مورد استفاده تعیین می‌کنیم . سپس با استفاده از عدد زمینه که به عنوان اندیس استفاده می‌شود و جدول احتمالات یك کد ریاضی را پیدا می‌کنیم. همانطوریکه قبلاً اشاره شد برای پیدا کردن عدد زمینه دو روش متفاوت وجود دارد که یکی مربوط به روش intra و دیگری مربوط به روش inter می‌باشد. در روش inter علاوخ بر استفاده از BAB که قرار است کد شوند از اطلاعات MC BAB نیز برای پیدا کردن عدد زمینه استفاده می‌شود. در روش intra که عدد ده بیتی مطابق با شکل زیر تولید می‌شود در حالیکه در روش inter با استفاده از اطلاعات MC BAB یك کد ۹ بیتی را تولید می‌کنیم.



شکل ۳-۱۳- نحوه پیدا کردن عدد زمینه در CAE

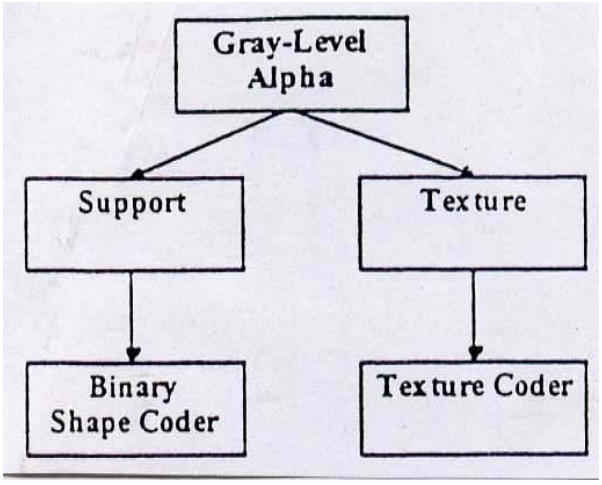
در روش بالا وقتی که عدد زمینه را محاسبه می‌کنیم ممکن است به اطلاعاتی خارج از BAB جاری نیاز داشته باشیم، در صورتیکه این اطلاعات در زمان محاسبه عدد زمینه در BAB هایی که همسایه BAB جاری هستند موجود باشند از این اطلاعات استفاده می شوددر صورتیکه نقاطی که به آن اشاره می‌شود خارج از VOP جاری باشد به جای این مقادی رصفر در نظر گرفته می‌شود و در صورتیکه در زمان محاسبه عدد زمینه این نقاط نامشخص باشند با استفاده از روش پر کردن مقدار این نقاط را مشخص می‌کنیم. با توجه به ترتیب کد کردن BAB ها در روش intra امکان دارد C2,C3,C7 ناشناخته باشند در حالیکه در روش inter امکان دارد C1 ناشناخته باشد. اگر در روش intra مقدار C7 نامشخص باشد در آن صورت

C7=C8 قرار می‌گیرد، اگر C3 نامشخص باشد C3=C4 قرار می‌گیرد ، اگر C2 نامشخص باشد C2=C3 قرار میگیرد و در inter اگر C1 نامشخص باشد C1=C2 قرار می‌گیرد . بعد از تولید کدهای ریاضی CAE به منظور اینکه از شبیه شدن این کدها با کد شروع یا (Start of Code) SC جلوگیری شود یکسری تمهیداتی در نظر گرفته می‌شود به این منظور اگر در ابتدا و یا وسط و آخر یك رشته داده تعداد معینی صفر قرار گرفته باشد واز حد تعیین شده که توسط اعداد MAX TALLING, MAX MIDDLE, MAX HEADING بیشتر باشد ، یك بیت یك بیت ؤ داخل رشته کد شده، اضافه می‌شود تا از شباهت آن با SC جلوگیری شود اشتباه نشود و در زمان واکد کردن یك های اضافه شده ، حذف خواهد شد.

۳-۳-۵- کد کردن نقشه خاکستری آلفا

به طور کلی اطلاعات نقشه خاکستری را میتوان به دو دسته تقسیم کرد. یك تابع که حاوی مقدارهای آلفا می‌باشد و مثل روشهای مورد استفاده در کد کردن بافت (Texture) که در ادامه به آن اشاره می شود کد می‌شود و قسمت دیگر به تابع حمایت (Support) معروف می‌باشد و همان نقشه دودوئی آلفا می‌باشد و با در نظر گرفتن آستانه صفر، از روی اطلاعات اولیه بدست آمده است و مانند روشهایی که در بالا ذکر شد کد می‌شوند.

اطلاعات کد شده ماکروبلوکهای آلفا بعد از کد شده به انتهای اطلاعات کد شده مربوط به بافت تصویر اضافه می‌شوند.



شکل ۳-۱۴- کد کردن نقشه خاکستری آلفا

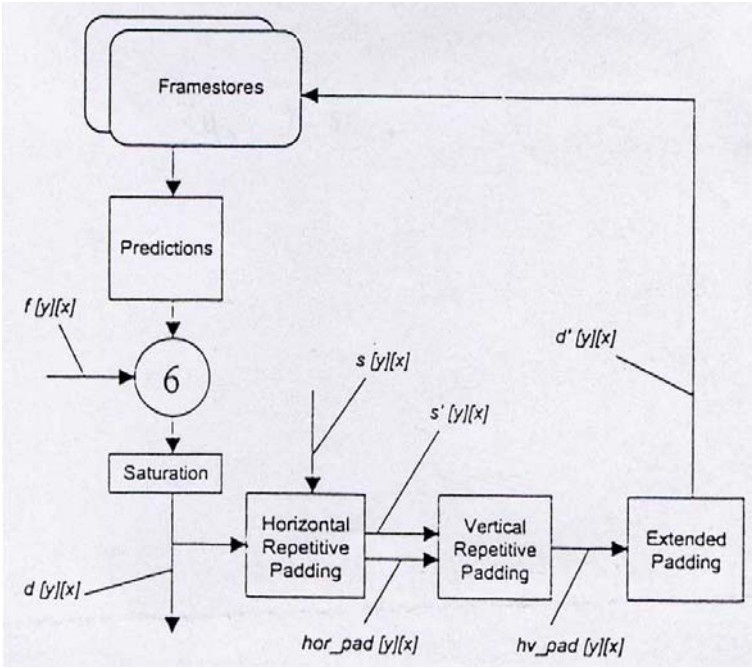
۳-۴-روشهای تخمین حرکت

با توجه به اینکه بر خلاف روشهای سابق در این استاندارد این امکان وجود د ارد که یك VOP یك شکل دلخواه داشته باشد بنابر این پیاده سازی روشهای مانند پیدا کردن بردار حرکت و

تخمین حرکت نیاز به یکسری پردازشهای اولیه دارد همانطوریکه قبلاً اشاره شد بعد از نقسیم یك VOP به ماکروبلوکها ۱۶×۱۶ برخی از ماکروبلوکها که روی مرز VOP قرار دارند حالت چند ضلعی (polyon) پیدا میکنند که قبل از پیدا کردن بردار حرکت بایستی ابتدا عملیات پر کردن (padding) روی VOP مرجع انجام شود برای این کار از یکسری روشهای تکرار شونده (Repetitive) مبتنی بر ماکروبلوك استفاده می‌شود . پر کردن ماکروبلوکها قبل از پیاده سازی روشهای جبران سازی حرکت اجباری است.

۳-۴-۱- پر کردن ماکروبلوکها چند ضلعی

شکل زیر نموداری از عملیات پردازشی پر کردن را نشان می‌دهد . این روش در دو مرحله افقی عمودی صورت می‌پذیرد. ماکروبلوکهایی که خارج از VOP واقع شده‌اند را با استفاده از روشهای بسط یافته (Extended) پر می‌کنیم. عملیات پر کردن بر این همه ماکروبلوکهایی که داخل چارچوب مرزی است باید صورت گیرد. ابعاد این چارچوب مرزی برای مولفه رنگ و روشنایی توسط Width-VOP و heightVOP مشخص می‌شود که در مولفه رنگ ضرب در ۸ و در مولفه روشنایی ضرب در ۱۶ می‌شود.



شکل ۳-۱۵- پیاده سازی مراحل پر کردن

مطابق شکل یك ماکروبلوك d[y][x] با توجه به اطلاعات متناظر شکلش که در S[y][x] موجود می‌باشد با استفاده از روشهای تکرار شونده افقی (Horizontal repetitive Padding) سپس روشهای تکر شونده عمئودی (Vertical repetitive padding) پر می‌شود.

در روش پر کردن افقی اگر در نقطه (x,y) اطلاعات d[y][x] موجود نباشد یعنی S[y][x] صفر باشد این نقطه با نقطه مجاورش پر میشود در نتیجه در صورتیکه در پیکسل در مجاورتش موجود باشد با میانگین این دو پیکسل پر میشود و نتیجه hor-pa d[y][x] گذاشته می‌شود و در صورتیکه این پیکسل مقدار داشته باشد مقدار موجود این نقطه در hor-pa d[y][x] گذاشته می شود الگوریتم پیاده سازی آن مطابق شکل زیر می‌باشد.

بعد از اینکه پر کردن افقی تمام شد نقاطی که در روش بالا پر نشده‌اند با استفاده از پر کردن عمودی پر می‌شوند که کد پیاده سازی آن نیز مطابق الگوریتم زیر می‌باشد.

الگوریتم پر کردن افقی

```
for(x=0;x<N;x++){
if (s[y][x]= =1){hor-pad[y][x]=d[y][x];S'[y][x]=1;}
else{
if (s[y][x']= =1& &S[y][x˝]= =1){
hor-pad[y][x]=(d[y][x']+d[y][x˝])//2;
s´[y][x]=1;
}ELSE IF (s[y][x´]= =1){
hor-pad[y][x]=D[y][x´];s´[y][x]=1;
}elseif (s[y][x˝]= =1){
hor-pad[y][x]=D[y][x˝];s´[y][x]=1;
}
}
```

در صورتیکه یك ماکروبلوك اصلاً داخل VOP نباشد ولی در چارچوب مرزی موجود باشد ودر همسایگی یکی از ماکروبلوکها روی مرز باشد در این صورت از روش بسط یافته برای پر کردن آن استفاده می‌شود. برای پر کردن طبق اولویت ارائه شده از یکی از ماکروبلوکهای پر شده که در همسایگی‌اش موجود است استفاده می‌کند و همه ماکروبلوك را با اطالعات روی لبه ماکروبلوك مجاور پر می‌کند. بالاترین اولویت مربوط به ماکروبلوك پایینی این ماکروبلوك می‌باشد . اولویت بعدی مربوط به ماکروبلوك سمت راست و اولویت سوم مربوط به ماکروبلوك سمت چپ اولویت چهارم مربوط به ماکروبلوك بالا می‌باشد. سایر ماکروبلوکهایی که در همسایگی یك ماکروبلوك مرزی واقع نشده است را با ۱۲۸ پر میکنیم.[5,6,7]

۳-۴-۲- انطباق ماکروبلوکهای چند ضلعی

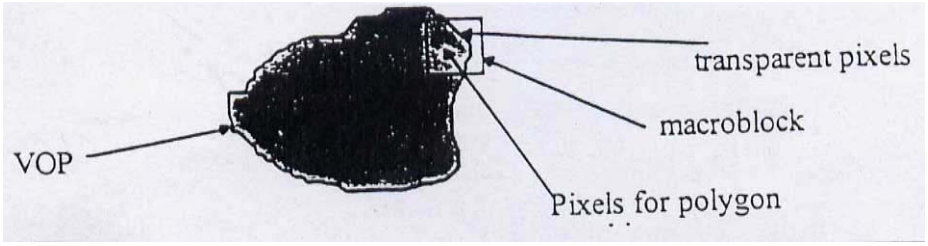
به منظور پیداکردن بهترین ماکروبلوك در VOP مرجع که انطباق مناسبی با یك ماکروبلوك در VOP جاری داشته باشد، ابتدا VOP مرجع را با استفاده از روشهای پر کردن که در بالا اراده شد پر می‌کنیم. البته این عملیات متناسب با اطلاعات شکل VOP مرجع صورت میگیرد یعنی مثلاً اگر VOP مرجع کوچکتر از VOP جاری باشد به اندازهVOP جاری بسط پیدا نمی کند بلکه مطابق آنچه در چارچوب مرزی خودش مشخص شده عملیات پر کردن صورت می‌پذیرد ابتدا VOP جاری را متناسب با طول وعرض چارچوب مرزی اشت و نقطه شروع این چارچوب تقسیم به بلوکهای ۱۶×۱۶ می‌کنیم و متناسب با اطلاعات نقشه آلفا که حضور یا عدم حضور یك نقطه داخل VOP را نشان می‌دهد. عملیات جستجو را آغاز می‌کنیم. البته مجموع

قـــــــــــــدر م طلـــــــــــــق

SAD(sum of Absolute Difference) را تنها برای نقاطی که داخل VOP جاری هستند محاسبه می‌کنیم . به عبارت دیگر SAD روی یك ماکروبلوك چند ضلعی که با بردار حرکت (x,y) در نظر گرفته می شود از روش زیر محاسبه می‌شود.

$$SAD_N(x,y)=\sum_{i=1,y=1}^{NN}|original-previos|*\left(!\left(Alpha_{original=0}\right)\right)$$

حداکثر ناحیه جستجو می‌تواند در محدوده [-64,63] باشد یعنی $x.y\in[-64,63]$ می‌تواند باشد در صورتیکه عملیات جستجو برای بلوکهای ۱۶×۱۶ صورت بگیرد N=16 و در غیر اینصورت N=8 در نظر گرفته می شود.



شکل ۳-۱۶- یك ماکرو بلوك با شکل دلخواه

$$SAD_{K\times 8}=\sum_{i=1}^{k}SAD_i(x,y) \qquad 0<K<=9$$

میزان حداقل SAD در یك روش پیش گویی خطا عبارتست از :

$SAD_{inter}=\min(SAD_{16}(x,y),SAD_{K*8})$

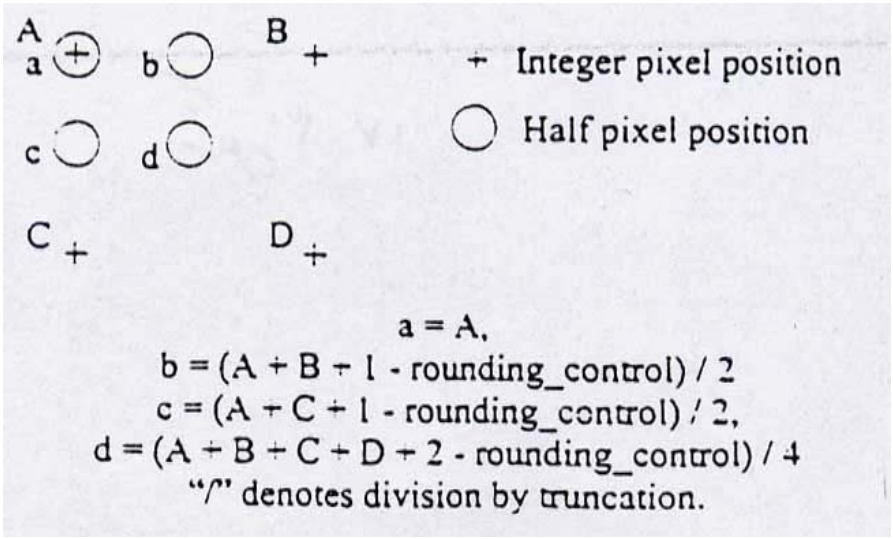
در این مرحله برای انتخاب روش کد کردن ماکروبلوك ابتدا میانگین همه مقدارهایVOP را محاسبه می‌کنیم و سپس تفاوت ماکروبلوك جاری از ماکروبلوکی که همه مقدارهای آن مساوی با مقدار میانگین باشد محاسبه می‌شود و بامقایسه این اختلاف با SAD_{inter} در مورد اینکه از چه روشی برای کد کردهن این ماکروبلوك استفاده شود تصمیم گیری می‌شود.

$$MB-mean=\left(\sum_{i=1,y=1}^{NC} original\right)/NC$$

$$A=\sum_{i=1,j=1}^{16,16}\left|original-(MB-mean)\right|*\left(!\left(Alpha_{original}=0\right)\right)$$

در روش بالا اگر که (A<(SAD $_{inter}$ –(2*NB)) روش کد کردن درون قابی (intra) انتخاب می‌شود، در غیر اینصورت این روش کد کردن بین قابی(inter) انتخاب می شو در حالیکه بخواهیم از روش کد کردن بین قابی(inter) استفاده کنیم بایستی مراحل پیدا کردن بردار حرکت را روی نصف پیکسل ادامه دهیم.

جستجو روی نصف پیکسل (Half Sample) هم روی بلوکهای ۱۶×۱۶ و هم بلوکهای ۸×۸ پیاده سازی می‌شود جستجوی نصف پیکسل در انتهای بردارهای V4,V3,V2,V1,V0 انجام می شود و روی فاصله ± 1 (نصف نمونه) جستجو صورت می‌پذیرد.

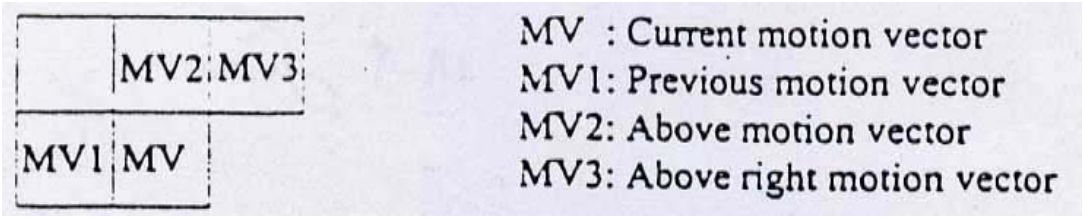


شکل ۳-۱۷- طرح جستجو برای پیاده سازی جستجوی نصف نمونه

بعد از پیدا کردن بهترین بردار حرکت در انتهای برادرهایی که مربوط به پیش گویی حرکت روی بلوکهای ۱۶×۱۶ و ۸×۸ می باشد. بهترین بردار حرکت پیدا شده را با نامصورت MV در نظر می گیریم که به صورت مولفه های (MVx,MVY) معرف می شوند. در این مرحله بایستی با مقایسه دو روش در نظر گرفتن یك ماکروبلوك به صورت یك بلوك ۱۶×۱۶ و یا تبدیل آن به چهار بلوك ۸×۸ یکی را انتخاب کنیم و به این منظور هر کدام از این روشهای که تولید SAD کمتری دارند به عنوان روش مورد استفاده انتخاب می شوند.

برای ارسال این برادرها و به منظور کاهش دادنه حجم داده ارسال شده از روشهای تفاضلی و با در نظر گرفتن بردارهای حرکت بلوکهای مجاور به عنوان تخمینی از این بردار حرکت استفاده

می شود مطابق شکل زیر MV3,MV2,MV1 به عنوان پیش گویی تقریب از MV که بایستی ارسال شود در نظر گرفته می شود.



شکل ۳-۱۸- پیش گویی بردار حرکت

به منظور پیش گویی بردار حرکت ، میانه مولفه‌های این بردارها روی جهت X به عنوان P_x میانه مولفه های این بردار ها روی جهت y به عنوان P_y انتخاب می شوند.

$P_x=Median(MV1_x,MV2_x ,MV3_x)$

$P_y=Median(MV1_y,MV2_y ,MV3_y)$

برای ارسال MV رابطه زیر استفاده می شود:

$MVD_x=MV_x -P_x$

$MVD_y=MV_y –P_y$

در صورتیکه یکی از بلوکهای بالا خارج از ناحیه VOP باشد بردار حرکت متناظرش صفر در نظر گرفته می شود. اگر دو عدد از این بلوکها خارج از VOP باشند بردار حرکت روی آنها مساوی با بردار حرکت سومین بلوك قرار می گیرد و در صورتیکه هر سه خارج از VOP باشد بردار حرکت هر سه مساوی با صفر در نظر گرفته می شود.

مفاهیم بالا اساس پیاده سازیهای روشهای مختلف پیدا کردن بردارهای حرکت را توضیح می دهد والبته چندین روش مختلف پیدا کردن بردارهای حرکت را توضیح می دهد و البته چندین

روش مختلف برای پیدا کردن بردار حرکت موجود می باشد که به صورت مختصر هر کدام شرح داده می شود.

۳-۴-۳- روشهای جبران سازی و تخمین حرکت به صورت نامحدود

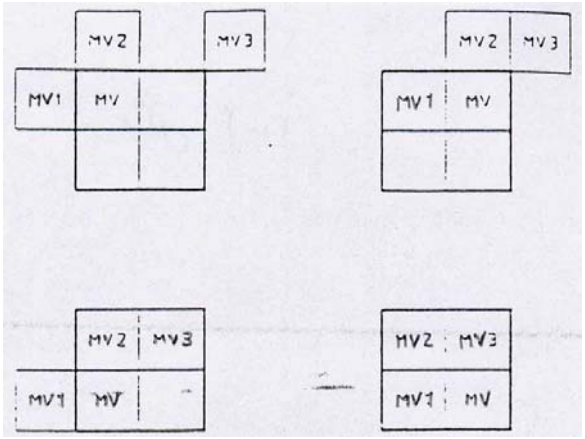
این روش برای بهبود بخشیدن به شرایط جستجوی بردار حرکت پیاده شده است. در روشهای متداول در صورتیکه بردار حرکت، اشاره به نقطه‌ای خارج از VOP پر شده مرجع بکند از این بردار حرکت صرفه نظر می شود در حالیکه در این روش VOP مرجع به اندازه کافی گسترش پیدا می کند و سپس پر می شود و عملیات جستجو روی آن صورت می گیرد. مطابق شکل زیر در صورتیکه نقطه ای که بردار حرکت به آن اشاره می کند جایی در خارج از VOP مرجع (منظور از VOP همان چار چوب مرزی اینVOP می باشد) باشد. در اینصورت از نمونه ای که در لبه چارچوب مرزی وجود دارد برای برگرداندن مقدار مشخص شده استفاده می شود.

مطابق شکل زیر Xref و yref مطابق معادلات زیر بدست می‌آیند.

Xref =MIN (MAX(xcurrdx ,vhmcsr),xdim+vhmcsr-1)

Yref =MIN (MAX(xcurrdx ,vhmcsr),xdim+vhmcsr-1)

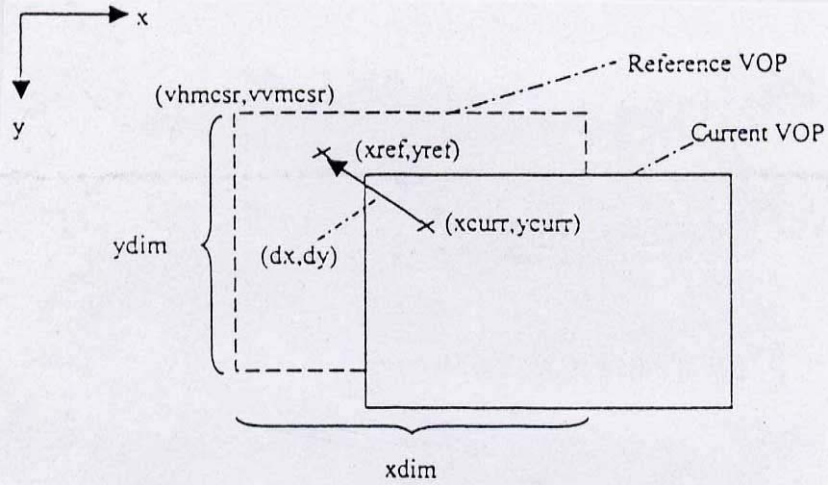
عملیات فوق روی بلوکهای رنگ نیز مشابه بلوکهای روشنایی پیاده سازی می شود.



شکل ۳-۱۹- روش نامحدود جبران سازی حرکت

۳-۴-۴- استفاده از روشهای پیش گویی پیشرفته

در صورتی که برای ارسال یك ماکروبلوك از چهار بردار حرکت استفاده شده باشد برای انتخاب بردارهای پیش گویی مطابق شکل زیر عمل می شود.

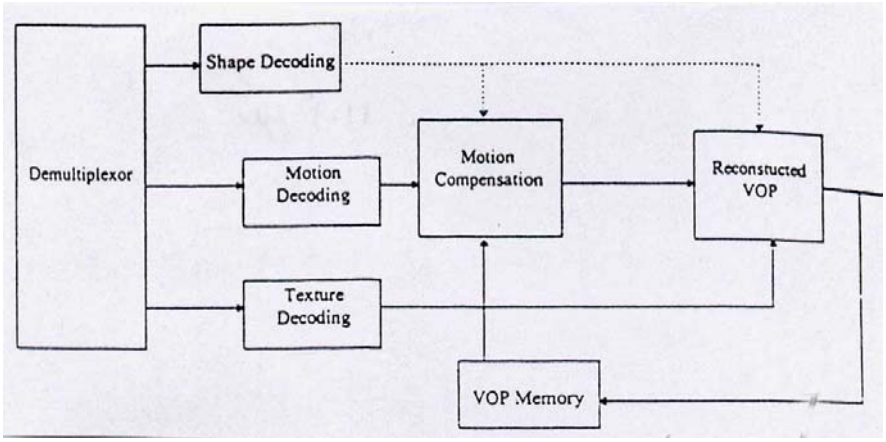


شکل ۳-۲۰- تعیین MV3,MV2,MV1 برای هر بلوك

بردار MVD که برای هر دو بلوك مربوط به رنگ مورد استفاده واقع می شود با در نظر گرفتن K بلوك داخل ماکروبلوك متناظر که هر کدام یك بردار حرکت را تولید می کنند و داخل VOP نیز واقع هستند بدست می آید.[8,9]

۳-۵-واکد کننده

شکل زیر نمای کلی از واکد کننده یك VOP را نشان میدهد. این واکد کننده به صورت عمده از دو بخش تشکیل شده است. در یك قسمتاطلاعات شکل باز سازی می شود و در قسمت دیگر واکد کننده با اطلاعات مربوط به بردارهای حرکت و بافت تصویر و اطلاعات شکل، تصویر نهایی را تولید میکند.



شکل ۳-۲۱-واکد کننده

۳-۵-۱- واکد کردن اطلاعات شکل

در صورتیکه=Video-Object-Layer-Shape" "∞ باشد در این صورت یك قاب مستطیلی مساوی با کل قاب تصویر در نظر گرفته شده است به همین دلیل نیازی به باز سازی و دریافت شکل نمی باشد ولی در غیر اینصورت با داشتن VOP-width و VOP-height و نقطه شروع چارچوب مرزی یك VOP را (توسط اطلاعات موجود در VOL منتقل می شود). می توان چارچوب مرزی VOP را تشکیل داد و با در نظر گرفتن مقدار VOP-Shape-Coding-Type تعیین می کنیم که روش کدینگ در این VOP از نوع I و یا P و یا B خواهد بود با داشتن مقدار "Chang-Conv-vaito-disable" می توانیم تشخیص دهیم که ایا این BAB دچار تغییر اندازه شده است یا نه و ضریب CR را از روی "Conv-ration" مشخص می کنیم .

با در نظر گرفتن اطلاعات بالا شروع به واکد کردنه اطلاعات با ینری شکل برای VOP جاری می کنیم همانطوریکه در مباحث مربوط به کد کننده اشره شد هر BAB می تواند یکی از هفت حالت زیر را داشته باشد.

Bab_type	Semantic	Used in
0	MVDs==0 && No Update	P- ,B-VOPs
1	MVDs!=0 && No Update	P- ,B-VOPs
2	Transparent	All VOP types
3	Opaque	All VOP types
4	IntraCAE	All VOP types
5	MVDs==0 && interCAE	P- ,B-VOPs
6	MVDs!=0 && interCAE	P- ,B-VOPs

جدول ۳-۱- لیست انواع مختلف BAB

در صورتیکه نوع VOP از نوع I باشد برای اینکه نوع کد VLC را معین کنیم از روی BAB های همسایه ضریبی به نام C را پیدا می کنیم که مطابق جداول ضمیمه در استاندارد MPEG-4 نوع VLC مورد استفاده مشخص می شود اگر فرض بگیرم f(i.j) نوع BAB موجود در ستون I ام و سطر j ام باشد مقدار C از رابطه زیر بدست می آید.

C=27* (f(i-1,j-1)-2)+9*((f(i+1,j-1)-2)+3*(f(i+1,j-1)-2)+(f(i-1, j)-2)

در صورتیکه BAB از نوع P یا B باشد نوع VLC وابسته به BAB که در موقعیت BAB جاری ولی در VOP مرجع قرار دارد می باشد در صورتیکه از نوع P باشد آخرینP یا I قبلی به عنوان مرجع در نظر گرفته می شود ودر صورتیکه از نوع B باشد و یکی از تزدیکترین P

یا I که خالی نباشند و نزدیکتر به این B باشد به عنوان مرجع انتخاب می شوند. البته این امکان وجود دارد که سایز VOP جاری و مرجع با یکدیگر متفاوت باشد و برای اینکه BAB متناظر پیدا شود بایستی ابعاد این چار چوب را کاهش داد ویا آن را بسط داد. اگر ابعاد VOP جاری از VOP مرجع کوچکتر باشد آخرین سطر و یا ستون از VOP مرجع حذف می شود. تا هم اندازه شوند و در صورتیکه VOP جاری بزرگتر باشد آخرین سطر و یا ستون آنقدر تکرار می شوند تا هر دو اندازه باشند.

در صورتیکه BAB جاری "Trans parent" و یا "Opaque" باشد با مقدارهای صفر یا ۲۵۵ پر می شد و نیازی به ادامة مراحل واکد کردن اطلاعات شکل نمی باشد در غیر این صورت بایستی عملیات جبران سازی خطا و سایر پردازشها روی آن صورت می گیرد.

در حالیکه نوع BAB جاری "Noupdate" باشد تنها عملیات جبران سازی حرکت برای آن انجام می پذیرد و دیگری نیاز به واکد کردن CAE نمی باشد.

در صورتیکه در یك BAB مقدار MVDs!=0 در این صورت اختلاف بردار حرکت در راستای x,y بایستی محاسبه شود، به این منظور با استفاده از جداول مربوط به کدهای VLC این مقدارها واکد می شوند. سپس با در نظر گرفتن MVPs همانطوریکه در قسمت کد کننده توضیح داده شد مقدار نهایی MVs پیدا می شود.

البته اگر VOP از نوع B باشد یا اطلاعات شکل از نوع دودوئی باشد فقط MVs- MVs_3, MVs_{1s},MVs_2 را در نظر می گیریم و در صورتیکه BAB از نوع "Transparent" باشد MVs برای این BAB در نظر گرفته نمی شود و سراغ اولویت بعدی می رود.

بعد از پیدا کردن بردار حرکت ، مرحله باز سازی BAB جاری می باشد در صورتیکه از نوع 1,0 باشد یك قسمت از اطلاعات دودوئی از روی P,I قبلی روی این BAB کپی می شود اما در صورتیکه از نوع ۴،۵ باشد نیاز دارد که عملیات دی کد کردن برای کدهای CAE صورت بگیرد . بدلیل اینکه نیاز داریم که ضرایبهای C را برای دی کد کردن CAE به دست بیاوریم برای هر BAB یك حاشیه معادل دو پیکسل در نظر می گیریم یعنی در نهایت این BAB ها ابعاد ۱۸×۱۸ دارد.

برای واکد کردن یك کد CAE بایستی چند مرحله را انجام دهیم ابتدا شبیه روشی که در کد کننده استفاده کردیم مقدار عدد C را برای هر دو حالت Intra, Inter محاسبه کنیم. سپس با داشتن مقدارC و از روی جداول موجود مقدار احتمال را محاسبه می کنیم. در مرحله آخر با توجه به مقدار احتمال محاسبه شده و الگوریتمهای ارائه شده برای واکد کردن CAE رشته اطلاعات را واکد کنیم.

در صورتیکه اطلاعات شکل از نوع نقشه خاکستری باشد همانطوریکه قبلاً اشاره شد اطلاعات آهن به دو دسته تقسیم می شود یك دسته مانند اطلاعات دودوئی کد واکد می شود و دستة دیگر شبیه اطلاعات بافت تصویر منتقل می شود با توجه به اینکه برای کد کردن اطلاعات آلفا ممکن است از روشهای پر کردن استفاده شده باشد در این صورت مقدارهایی که روی اطلاعات باینری صفر هستند مساوی با صفر در نظر گرفته می شوند. وقتی یك ماکروبلوك سطوح خاکستری را واکد می‌کنیم توسط مقداری که در "CODA" وجود دارد به نوع کد کردن اطلاعات این بلوك پی می بریم باید به این نکته توجه کرد واکد کردن اطلاعات بافت و اطلاعات سطوح خاکستری کاملاً مستقل از هم اتفاق می افتد یعنی ممکن است که اطلاعات بافت شکل موجود نباشد در حالیکه اطلاعات آلفا وجود داشته باشد.

در صورتیکه "CODA=all opaque" (در مورد B.P.I) و یا اینکه "CODA=not coded" باشد (در مورد P یا B) در این صورت دیگر اطلاعات مربوط به آلفا وجود نخواهد داشت.

در غیر اینصورت ابتدا اطلاعات cbpa(coded block pattern) منتقل می شود و بعد از آن اطلاعات بافت کشل روی بلوکهای ۸×۸ منتقل می شود و البته تنها آن بلوکهایی که داخل VOP هستند کد می شوند.

در صورتیکه مقدار QP روی ماکرو بلوکهای شکل ، ثابت باشد نیازی به مطابق روز کردن QP نخواهد بود اما در صورتیکه مقدار QP متغیر باشد در ابتدای واکد کردن هر ماکروبلوك بایستی مقدار جاری این QP را با استفاده از مقدار dquant(Different Quantisation) به دست آورد.

در صورتیکه اطلاعات روشنایی یك ماکروبلوك به صورت درون قابی کد شده باشد حتماً اطلاعات آلفا نیز به صورت درون قابی کد می شود برای کد کردن ضریب DC از همان روشهایی که برای مقیاس پذیر کردن ضریب DC اطلاعات روشنایی استفاده شد استفاده می شود به همین دلیل از روی "alpha -quant" بایستی مقدار "DCS colorA" را پیدا کنیم. پیش گوی روی ضرایب DC نیز مانند روشی که برای اطلاعات روشنایی استفاده شد به کار میرود ولی با توجه به اینکه ممکن است حالیت اتفاق بیفتد که یك ماکروبلوك اصلاً کد نشود برای پیش گویی سایر ماکروبلوکها یك مقدار مشخص را از روی ضرایب مقدارهای تعریف شده طبق لیك فرمول از قبل مشخص ، محاسبه کرده به عنوان ضریب DC متناظر با این ماکروبلوك که کد نشده است برای پیش گویی استفاده می کنیم.

در صورتیکه از B-VOP, P-VOP استفاده کنیم برای کد کردن اطلاعات آلفا نظیر اطلاعات روشنایی از همان روشهای پیش گویی حرکت استفاده می شود.

همانطوریکه قبلاً اشاره شد اطلاعات آلفا مقداری از اطلاعات روشنایی را در خود حمل می کند و برای یکنواخت کردن لبه شکل به کار می رود پس از واکد کردن اطلاعات روشنایی و اطلاعات آلفا ، این دو مقدار بایستی با یکدیگر ترکیب شوند وبه شکل نهایی را تولید کنند.
نحوه ترتیب اطلاعات دو تا لایه تصویر با استفاده از اطلاعات آلفا بستگی به نوع روش توصیف شده برای ترکیب کردن دارد به عنوان مثال در صورتیکه ″Composition-method=0″ باشد روش محو کردن لبه ها یا (cross-fading) استفاده می شود و اگر لایه N با n بیت اطلاعات آلفا روی لایة M قرار می گیرد تا اطلاعات نهایی به صورت لایة P تولید شود در این صورت خواهیم داشت که:

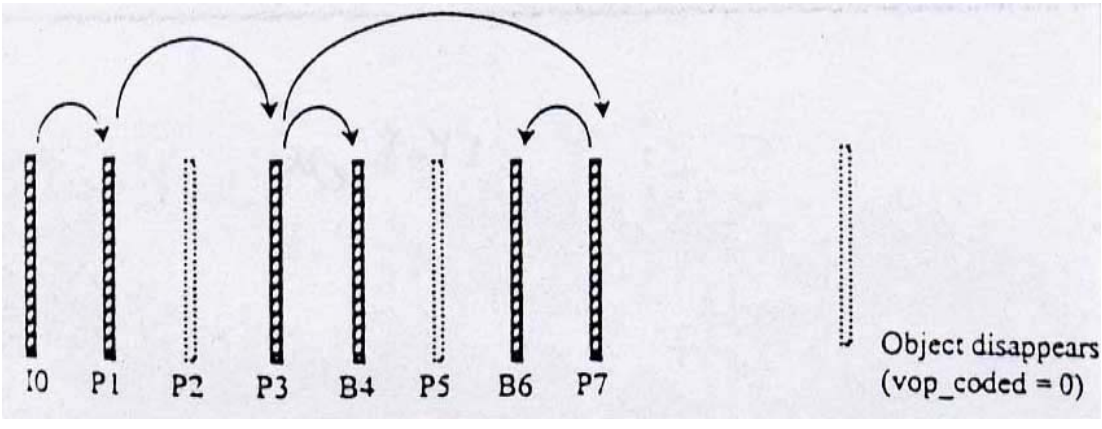
P_{YUV} =((2″-1-Nalpha)* M_{YUV} +(Nalpha *N_{YUV}))/(2″-1)

Palppha = (2″-1)

در صورتیکه ″Composition-method=1″ باشد روش ترکیب جمع شونده با (Additire mixing) استفاده می شود و اگر یك لایه N با n بیت اطلاعات آلفا را روی لایه M قرار دهیم تا تولید لایه P بکند مقدارهای تولید شده در لایه P مطابق رابطه زیر بدست خواهد آمد که در رابطه زیر ″Black″ مسطح روشنایی مربوط به رنگ سیاه را در تصویر مشخص می کند.

۳-۵-۲- پیش گویی زمان اطلاعات

این امکان در این استاندارد وجود دارد که هنگام پیش گویی زمان اطلاعات برخی از قابهای تصویر حذف می شود در صورتیکه چنین اتفاقی بیفتد برای یافتن قاب مرجع پیش گویی در برخی موارد ممکن است دچار مشکل شویم. برای حل این مشکل مطابق شکل زیر در صورتیکه یك قاب از نوع p ، نزدیکترین مرجعش حذف شده باشد از نزدیکترین I و یا P که قبل از آن واکد شده است استفاده می شود و در صورتیکه قاب از نوع B یکی از مراجع پیش رو و یا برگشتی اش حذف شده باشد از مرجع باقی مانده برای پیش گویی استفاده می شود و در صورتیکه هر دو مرجع حذف شده باشند مقدار ماکروبلوکهای مرجع در اطلاعات روشنایی و رنگ و اطلاعات آلفا با مقدارهای از پیش تعیین شده ای پر می شود.



شکل ۳-۲۲- ساختار پیش گویی زمانی در صورت ناپدید شدن یك قاب

۳-۵-۳- مفاهیم اصلی واکد کردن اطلاعات مقیاس پذیر

همانطوریکه در قسمتهای مربوط به کد کردن یك رشته اطلاعات به صورت مقیاس پذیر ذکر شد. یك VOP در لایه به ساخت می تواند به یکی از دو صورت p_vop و یا B-POV کد شود. اینکه در هر کدام از این روشها قابهای مرجه چگونه انتخاب شده اند توسط یك پرچم (flag) در واکد کننده منتقل می شود که مطابق جداول خاصی قابل شناسیایی می باشد وابسته به نوع مقیاس پذیری ، مرجع پیش رو ، یا بازگشتی می تواند یکی از vop های داخلی لایه پایه و یا لایه به ساخت باشد. در مورد مقیاس پذیری مکانی که از قاب تصویر در همان زمان به عنوان مرجع پیشگویی استفاده می شود دیگر نیازی به ارسال بردارهای حرکت نمی باشد، البته می توان با در نظر گرفتن B-VOP هم به صورت زمانی و هم به صورت مکانی عملیات پیش گویی را انجام داد که در اینصورت بردارهای حرکت نیز در نظر گرفته می شوند.

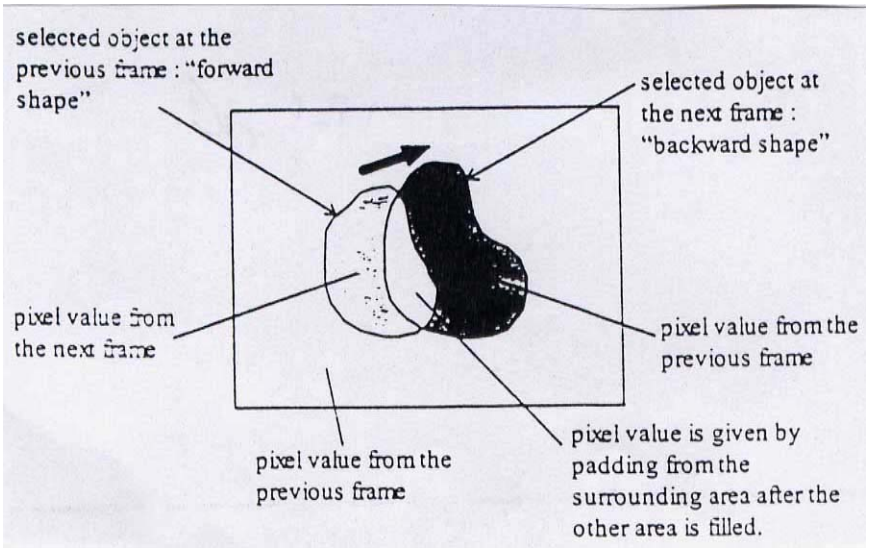
در واقع مفهوم مقیاس پذیری مکانی به این معنی است که قدرت تفکیك پذیری مکانی تصویر پایین می باشد به عنوان مثال در صورتیکه لایة پایه دارای تفکیك پذیری QCIF باشد لایه بهساخت تولید یك تصویر با قدرت تفکیك پذیری CIF می شود . برای واکد کردن این اطلاعات ابتدا لایه پایه به صورت معمولی واکد می شود سپس روی اطلاعات پایه با استفاده از روشهایی که قبلاً ذکر شد عملیات افزایش نمونه صورت می گیرد و به عنوان مرجع پیش گویی در واکد کردن اطلاعات مربوط به لایه بهساخت استفاده می شود.

یك VOP از نوع B در لایه بهساخت بیشتر عملکردی شبیه به یك P-VOP دارد از آنجاییکه می توان به صورت مرجع پیش گویی برای یك P و یا B دیگر در لایه بهساخت مورد استفاده واقع شود. در صورتیکه VOP در لایه بهساخت از نوع P-VOP باشد مرجع پیش گویی برای این I-VOP:VOP متناظر زمانی با آن در لایه پایه می باشد، در این روش برای کاهش دادن سر رشته های اطلاعات بردارهای حرکت را صفر در نظر گرفته اند.

اگر VOP در لایه بهساخت از نوع B-VOP باشد متناسب با نوع قابهای انتخاب شده به عنوان مرجع پیش رو برگشتی این اطلاعات واکد می شود . در صورتیکه یك قاب از نوع P در لایه

پایه به عنوان مرجع پیش گویی در نظر گرفته شده باشد و در همان لحظه زمانی باشد در اینصورت بردار حرکت را صفر در نظر می گیریم.

در صورتیکه از مقیاس پذیری زمانی استفاده کنیم و نوع مقیاس پذیر کردنصحنه از یك نوع باشد بایستی پس زمینه را با استفاده از روشهای ترکیب پس زمینه (Back ground Composition) و با استفاده از دو دسته اطلاعات مربوط به شکل در دو قاب متوالی تولید کنیم شکل زیر مفهوم ترکیب پشت صحنه را واضح تر نشان می دهد.



شکل ۳-۲۳- ترکیب پشت صحنه

۳-۵-۴- تولید صحنه نهایی

همانطوریکه اشاره شد خروجی هاییك واکد کننده در این روش عبارتند از VOP های مختلف که به منظور تولید صحنه نهایی بایستی این VOP ها را با یکدیگر ترکیب کنیم برای مثال اگر فرض بگیریم یك VOP دو لایه N را بر روی VOP در لایه M می گذاریم تا لایه P را تولید کنیم مقدارهای نهایی مساوی Y,U,V تولید شده در لایه P به صورت زیر بدست خواهد آمد.

$$P_{YUV} = ((255-N_{alpha})*M_{YUV} + (N_{alpha}*N_{YUV}))//255$$

$$P_{alpha} = 255$$

اگر بیش از دو VOP در یك صحنه موجود باشد با در نظر گرفتن تصویری که از ترکیب بالا تولید شده باشد به عنوان پشت صحنه عملیات ترکیب را یکبار دیگر انجام می دهیم.در این فصل به صورت اجمالی مهمترین روشهای کد و واکد اطلاعات روی اشیاء ویدئویی را بررسی کردیم در ادامه فلوچارت ساده ای از کد کننده و واکد کننده ارائه می شود.[6,7,8,9,10]